


```
163 CCGAAG.....GATATTATAGACTCGTTAAAGGGGGCATCGGAG 203
   : : : : : : : : : : : : : : : : : : : : : : : : : :
59 GluArgAsnArgProGluTyrIleAspLeuLeuLysSerLeuIleLysly 75
   : : : : : : : : : : : : : : : : : : : : : : : : : :
204 TGACCTGATAGATAATCGAAGAGCTACAGCGACGCAATACTCCGCC 253
   : : : : : : : : : : : : : : : : : : : : : : : : : :
75 sGlyGlnValGluLeuValAlaGlyPheTyrGluProIleLeuValA 92
   : : : : : : : : : : : : : : : : : : : : : : : : : :
254 TCCTCCGCTTAGCAGAGTAGAAGACAAAGTTTCAGAGATAGGAAGTT 303
   : : : : : : : : : : : : : : : : : : : : : : : : : :
92 IalIlePro.....GluGluAsp 97
   : : : : : : : : : : : : : : : : : : : : : : : : : :
304 AAGGAAGAGCTCTCGAGCTTCT..... 327
   : : : : : : : : : : : : : : : : : : : : : : : : : :
98 ArgValGluGlnIleLysLeuSerLysGlyTrpAlaArgLysMetGlyTy 114
   : : : : : : : : : : : : : : : : : : : : : : : : : :
328 ....CQAAGGATTCGCTGCCAGAGCTCGCTATGACCCGATAATCC 373
   : : : : : : : : : : : : : : : : : : : : : : : : : :
114 rGluAlaArgGlyLeuTrpLeuThrGluArgValTrpGluProGluLeuV 131
   : : : : : : : : : : : : : : : : : : : : : : : : : :
374 CTGCCATCTAGAGCAACAGGTATGAGTATCTATCCGCGACGGGAG 423
   : : : : : : : : : : : : : : : : : : : : : : : : : :
131 allYsThrLeuArgGluAlaGlyIleGlnTyrValIleLeuAsp..... 145
   : : : : : : : : : : : : : : : : : : : : : : : : : :
424 CGGATGCTTTCTCAGCTCATCTCACTCGCG...ATAAGCCAATTAA 470
   : : : : : : : : : : : : : : : : : : : : : : : : : :
146 .....AspTyrHisPheMetSerAlaGlyLeuSerLysGluG 158
   : : : : : : : : : : : : : : : : : : : : : : : : : :
471 ACCGCTATCCACACTTATAAGGCCCAAGGAAAGCGCTTTAGGT 520
   : : : : : : : : : : : : : : : : : : : : : : : : : :
158 uLeuPheTrpProTyrThrGluAsnGlyGlyGluAlaIleValValP 175
   : : : : : : : : : : : : : : : : : : : : : : : : : :
521 AC.....ATCAGTATCTCTGCTCTCAGGGAGCTTAGG 555
   : : : : : : : : : : : : : : : : : : : : : : : : : :
175 heProIleAspGluLysLeuArgTyrLeuIleProPheArgProValasn 191
   : : : : : : : : : : : : : : : : : : : : : : : : : :
556 AAGCGGATAAGCTCGCTTT.....GAAGGTAAAGT 587
   : : : : : : : : : : : : : : : : : : : : : : : : : :
192 GluThrLeuGluTyrLeuHisSerLeuAlaAspGluAspGluSerLysVa 208
   : : : : : : : : : : : : : : : : : : : : : : : : : :
588 AAGCGTAAGGCAGCTCAAGACATCGAAGCGGTACCGTTGGTGGCGCG 637
   : : : : : : : : : : : : : : : : : : : : : : : : : :
208 lAlaVal....PheHisAspAspGlyGlyLysPheGlyAlaTrpProGly 224
   : : : : : : : : : : : : : : : : : : : : : : : : : :
638 TGAACACCGCTGTAATGCTCGGCATCGGAAGGCTTCTCTTATGAATCT 687
   : : : : : : : : : : : : : : : : : : : : : : : : : :
224 hrHisGluLeuValTyr..... 229
   : : : : : : : : : : : : : : : : : : : : : : : : : :
688 AAGAAAGTGGCGAGCTGATA.....GA 710
   : : : : : : : : : : : : : : : : : : : : : : : : : :
230 .....GluArgGlyTrpLeuLysGluPhePheAspArgIleSerSerAs 244
   : : : : : : : : : : : : : : : : : : : : : : : : : :
711 GGAAGAGACACATCTCTTATACGCGACCGATATAGATTATGCGC. 759
   : : : : : : : : : : : : : : : : : : : : : : : : : :
244 pasPlysIleAsnLeuMetLeuTyrSer.....GluTyrLeuSerL 258
   : : : : : : : : : : : : : : : : : : : : : : : : : :
760 .....TATAGGACATTGACGGCTAC...AGA 783
   : : : : : : : : : : : : : : : : : : : : : : : : : :
258 ysPheArgProLysLeuValTyrLeuProIleAlaSerTyrPheGlu 274
   : : : : : : : : : : : : : : : : : : : : : : : : : :
784 ATGAGTGTAGGGGATA.....TTAGAGGT 809
   : : : : : : : : : : : : : : : : : : : : : : : : : :
275 MetSerGluTrpSerLeuProAlaArgGlnAlaLysLeuPhePheGluPh 291
   : : : : : : : : : : : : : : : : : : : : : : : : : :
810 TATAGACAGCTCACTCGGAGCTGTGCTTCCCTCAGAGCTGAGGCACA 859
   : : : : : : : : : : : : : : : : : : : : : : : : : :
291 eileLysLysLeu...LysGluLeuAsnLeuPheGluLysTyrArg... 305
   : : : : : : : : : : : : : : : : : : : : : : : : : :
860 GTGGAAGGAGCTCTACTAGGACTTCAGTTCGGTGGCCACCATAGAGC 909
   : : : : : : : : : : : : : : : : : : : : : : : : : :
306 .....IlePheValArgGlyGlyIleTrp.....LysAsn 315
   : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
910 TTGAGGATATGGAGAGACGACGAAGGAAC.....GCAAGACTTAA 950
   : : : : : : : : : : : : : : : : : : : : : : : : : :
316 Phe...LeuTyrLysTyrProGluGlyAsnTyrMethHisLysArgMetLe 331
   : : : : : : : : : : : : : : : : : : : : : : : : : :
951 TATGCTGTCTCAATATATGAGGGGGAACTCGCCCTTTAGCCGAGAACA 1000
   : : : : : : : : : : : : : : : : : : : : : : : : : :
331 uMetLeuSer.....LysLeuLeuArgAsnAsn. 340
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1001 GCATGCAAGGGGATGGAGGCCCTCCCTGACAGGAGGCTGATGCCTTC 1050
   : : : : : : : : : : : : : : : : : : : : : : : : : :
341 .....ProThrAlaArgIlePheValLeu 348
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1051 CGGCGCATATATAACGAT.....TGGAGGGGTGAAATGGG 1086
   : : : : : : : : : : : : : : : : : : : : : : : : : :
349 ArgAlaGlnCysAsnAspAlaTyrTrpHisGlyValPheGly 362
   : : : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: SwissProt_40:AMY1_DICTH

seq_documentation_block:
ID AMY1_DICTH STANDARD; PRT; 685 AA.
AC P09361;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Alpha-amylase 1 (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase).
GN AMYA.
OS Dictyoglomus thermophilum.
OC Bacteria; Dictyoglomus group; Dictyoglomus.
OX NCBI_TaxID=14;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H-6-12;
RX MEDLINE=88225097; PubMed=2453362;
RT Fukusumi S., Kamizono A., Horinouchi S., Beppu T.:
RT "Cloning and nucleotide sequence of a heat-stable amylase gene from
RT an anaerobic thermophile, Dictyoglomus thermophilum.";
RL Eur. J. Biochem. 174:15-21(1988).
CC -!- FUNCTION: THIS AMYLASE IS A HIGHLY LIQUEFYING-TYPE: OLIGOMERS
CC APPEARED AT THE BEGINNING OF INCUBATION, FOLLOWED BY A GRADED
CC DECREASE IN THE AMOUNTS OF MALTOPIROSE, MALTOSE AND GLUCOSE IN
CC PROLONGED INCUBATION. IT IS HIGHLY HEAT-STABLE: THE OPTIMUM
CC TEMPERATURE FOR ITS ACTIVITY WAS FOUND TO BE AROUND 90 DEGREES
CC CELSIUS, BUT DECREASED TO ONLY 70% ACTIVITY AFTER 1 HOUR. NO
CC DECREASE OF ACTIVITY WAS OBSERVED WITHIN THE SAME TIME AT 80
CC DEGREES CELSIUS.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
```

```
alignment_scores:
  Quality: 157.50      Length: 315
  Ratio: 0.960        Gaps: 15
  Percent Similarity: 52.063  Percent Identity: 22.540
```

```
alignment_block:
US-09-886-400-3 x AMYL_DICTH ..
Align seg 1/1 to: AMYL_DICTH from: 1 to: 685

22 GCAACCTCCAGTATCGCGAATCCCAAGAGCGAAATCCCAAGAGTCAT 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 GlyAsnPheAspPheVal.....IleGluArgAlaIty 27

72 AGAAGAGCATACATCCCGAGTCATCGACACACTGATTAAGAA...GAAA 118
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27 rGluMetSerTyrLysProLeuIleAsnPhePhePheLysProAsp 44

119 TTCTTTGGGCTCAACATAACGGGCTATACCTTAAAGTTCTCTCCGAAG 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
44 heProIleAsnValHisPheSerGlyPheLeuLeuLeuTrpLeuGluLys 60

169 .....GATATTATAGACTCTTAAAGGGCGCATCGCGAGTGACCT 209
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 AsnHisProGluTyrPheGluLysLeuLysIleMetAlaGluArgGly 77

210 GATAGAGATAATCGAAGAGAGCTACACGACGCAATATCTCCCTCTCTCC 259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77 nIleGluPheValSerGlyPheTyrGluProIleLeuProIleLeuP 94

260 CGCTTAGCAGATAGACACACAGCTTACAGAGATAGGGAAGTTAAGAA 309
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
94 roAspLysAspLysValGlnGlnIleLysLysLeuAsnLysTyrIle 110

310 GAGCTCTCGAGCTTCTCCAAAGGATTCGTGCTCGCAGAGCTCGGCTA 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
111 AspLysPheGlyGlnThrProLysGlyMetTrpLeuAlaGluArgVal 127

360 TGACCCGATAATCCCTGCGCATCTACTGAAGACACACCGTTATGATGAT 409
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
127 pGluProHisLeuValLysTyrIleAlaGluAlaGlyIleGluTyrVal 144

410 TCGCCAGCGGGAGGCGATGCTTTCTCAGCTCATCTCAACTCGCGGATA 459
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
144 alValAsp...AspAlaHisPheSerValGlyLeu..... 155

460 AAGCCCAATTAACCGCTCTATCCACACCTTATAAAGGCCCAAGGAAA 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
156 ..... 159

510 GCCTTTAGTACATCAGCTATCTCTTGGTCTCAGGAGCTTAGGAGG 559
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
159 pLeuPheGlyTyr.....TyrLeuMetGluGluGlnGlyTyrLysLeu 174

560 CGATAAAGCTCGTTTGAAGTAAGGTAAAGCTAAAGCGACGTCAAGAC 609
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 la.....ValPheProIleSerMetLysLeuArgTyrLeu..... 185

610 ATCGAAGCCGCTACCCGTTGGTGGCGGTGAACACGGCTGTAATGCTCG 659
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 ..... 185

660 CATCGGAAGGCTTCTCTTATGATCTTAAAGAGTGGCGAGCTGGATA. 708
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 .....IleProPheAlaAspProGluGlnThrIleThrTyrLeu 199

709 .....GAGGACAGGACAACTTCTTATACGGCACCAGAT 744
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
199 spLysPheAlaSerGluAspLysSerLysIleAlaLeuLeuPheAsp 215

745 ATAGAGTTCATTGGC.....TATAGGACATTGCAGGCTA 779
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216 GlyGluLysPheGlyLeuTrpProAspThrTyrArgThrVal.....Tyr 230

780 CAGAGTGAAGTGTGAGGATATTATAGGTT...ATAGACGAGCTCAACT 826
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
230 rGlu.....GluGlyTrpLeuGluThrPheValSerLysIleLysG 244
```

```
827 CGGAAGTGTGCTTCCCTCA.....GAG 849
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 LuasPheLeuLeuValThrProValAsnLeuTyrThrTyrMetGlnArg 260

850 CTGAAGCAGCAGTGGAGGAGCTCTACTTACGAGCTTCGAGTTGG 894
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
261 ValLysProLysGlyArg...IleTyrLeuProThrAlaSerTyr 274

seq_name: SwissProt_40:AMYA_PYRAB

seq_documentation_block:
ID AMYA_PYRAB STANDARD; PRT; 655 AA.
AC Q9V298; 2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-amylase (EC 3.2.1.1).
GN AMYA OR PAB0118.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RC STRAIN=ORSAY;
SEQUENCE FROM N.A.
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC -|- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
linkages in oligosaccharides and polysaccharides.
CC -|- PATHWAY: POLYSACCHARIDE DEGRADATION.
CC -|- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ248283; CAB49100.1;
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57; 1.
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Complete proteome.
SQ SEQUENCE 655 AA; 77296 MW; 7F6F920B1A00EECE CRC64;

alignment_scores:
Quality: 145.50 Length: 401
Ratio: 0.808 Gaps: 22
Percent Similarity: 44.888 Percent Identity: 22.444

alignment_block:
US-09-886-400-3 x AMYA_PYRAB ..
Align seg 1/1 to: AMYA_PYRAB from: 1 to: 655

73 GAGAAGGCATACATCCAGTCATCGACACACTTAAAGAGAGAAATTC 122
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28 GluLysAlaTyrArgProPheLeuGluIleLeu.....GluGluTyrPr 42

123 T.....TTTGGCTCAACATAACGGGCTATACCTTAAAGTTCTCTCC 163
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
42 oAsnMetLysValAlaIleHisIleSerGlyIleLeuValGluTrpLeuG 59

164 CGAAG.....GATATTATAGACTCTTAAAGGGCGCATCGCGAGT 204
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 LuGluAsnLysProAspTyrIleAspLeuLysSerLeuValArgLys 75

205 GACCTGTATAGATATATCGAAGCAGCTACACGACGCAATCTCCCTCC 254
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 GlyGlnValGluIleValValAlaGlyPheTyrGluProValLeuAla 92
```

```
255 CCTCCCTTAGCAGATAGAACCAACAGTTTCAGAGATAGAGGAGTTA 304
    |||||
92 aileProLysGluAspArgLeuGluGlnIleTyrLeuLeuLysGluTrpA 109
    |||||
305 AGGAAGAGCTTCGAGCTTCCTCCAAAGGAGTTCGTGCTCCAGAGCTC 354
    |||||
109 laLysLysIle...GlyTyrAspAlaLysGlyLeuTrpLeuThrGluArg 124
    |||||
355 GCCATGACCCGATATCCCTGCCATCTACTAGGAGCAACGGTATGAGTA 404
    |||||
125 ValTrpGlnProGluLeuValThrLeuArgGluAlaGlyIleGluTyr 141
    |||||
405 TCTATTCCGCGAGGGGAGCGATGCTTTCTCAGCTCATCTCAACTCGG 454
    |||||
141 rValValValAsp.....AspTyrHisPheMetSerA 152
    |||||
455 CG...ATAAGCCAAATTAACCGCTCTATCCACACCTTATAAGGCCAA 501
    |||||
152 laGlyLeuSerLysAspGlnLeuPheTrpProTyrTyrThrGluAspGly 168
    |||||
502 AGGGA.....AGCGCTTAGGTAC..... 522
    |||||
169 GlyGluValIleThrValPheProIleAspGluLysLeuArgTyrLeuI 185
    |||||
523 .....ATCAGCTATCTCTGTGCTCTCAGGG 547
    |||||
185 eProPheArgProValAspLysValIleSerTyrLeuHisSerLeuAlaAs 202
    |||||
548 AGCTTAGGAGGCATTAAGCTGCTTTTCAAGGTAAGTAAGCTAAAG 597
    |||||
202 erGluAspGluSerLysValAlaValPheHis..... 212
    |||||
598 GCAGTCAAGACATCGAAGCGGTACCGCTTTGGTGGCGCGTGAACAGGC 647
    |||||
213 .....AspAspGlyGluLysPheGlyIleTrp..... 221
    |||||
648 TGTATGCTCGGCATCGAAGGCTTCTCTTATGAATCTTAAAGAGTGG 697
    |||||
222 .....ProMet.....Thr 225
    |||||
698 CGAGCTGATAGAGGACAAAGGAC..... 720
    |||||
225 yrGluTrpValTyrGluLysGlyTrpLeuArgGluPheAspArgVal 241
    |||||
721 .....ACATCTCTTATAGCGCACCATATAG... 750
    |||||
242 SerSerAspGluAlaIleAsnIleMetLeuTyrSerGluTyrLeuGlnLy 258
    |||||
751 .....TTCATTGGCTATAGGACATTGCGAGGTAC..... 780
    |||||
258 sPheLysProLysGlyLeuValTyrLeuProIleAlaSerTyrPheGluM 275
    |||||
781 .....AGATGAGTGTGAGGA 798
    |||||
275 etSerGluTrpSerLeuProAlaGlnGlnAlaLysLeuPheValGlu... 290
    |||||
799 TTATTAGAGGTATAGACGAGCTCAACTCGGAAGTGTGCTTCCCTCAGA 848
    |||||
291 PheValGluLysLeuLysGluLeuAsn..... 299
    |||||
849 GGTGAGCACAGTGAAGGAGGTCTTACTTACGAGCTTCGAGTTGGGCAC 898
    |||||
300 ....MetPheGluArgTyrArgValPheValArgGlyGlyIleTrp.... 313
    |||||
899 CAGATAAGAGCTTGAGATATGGAGAGGACGAAGGGRAC..... 939
    |||||
314 .....LysAsnPhe...PheTyrIleTyrProGluAlaAsnTyrMethHis 327
    |||||
940 GCAGACTTAATATGCTCTCTCAATATGAGGGCGCAACTCGCCCTTTT 989
    |||||
328 LysArgMetLeuMetLeuSer.....ArgLeuLe 337
```

```
990 AGCGGAGAACAGCGATGCAAGGGATGGAGCCCTCCCTGAGAGGAGGC 1039
    |||||
337 urGaspAsnProSerAlaArgPhe..... 346
    |||||
1040 TGGATCGCTTCCGGGCGATATATAAGAT.....TGGAGGGGTGAAAT 1083
    |||||
347 .....ValLeuArgAlaGlnCysAsnAspAlaTyrTrpHisGlyValPhe 361
    |||||
1084 GGG 1086
    |||||
362 Gly 362

seq_name: SwissProt_40:AMYA_METJA

seq_documentation_block:
ID AMYA_METJA STANDARD; PRT; 467 AA.
AC Q59006;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative alpha-amylase (EC 3.2.1.1).
GN MJ1611.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Adams M.D., Reich C.I.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Merrick J.M., Glodek A.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Fuhmann J.L., Nguyen D.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Sadow P.W., Hanna M.C.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii." 1996.
RL Science 273:1058-1073(1996).
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -I- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U67601; AAB99631.1; -.
CC TIGR: MJ1611; -.
CC InterPro: IPR004300; Glyco_hydro_57.
CC Pfam: PF03065; Glyco_hydro_57; 1.
CC KW Hypothetical protein; Hydrolase; Glycosidase; Carbohydrate metabolism;
CC Complete proteome.
CC SEQUENCE 467 AA; 55558 MW; 40A6B1CDD4D967E CRC64;
```

alignment_scores:
Quality: 141.50 Length: 377
Ratio: 0.741 Gaps: 24
Percent Similarity: 50.663 Percent Identity: 22.812

alignment_block:

US-09-886-400-3 x AMYA_METJA ..

Align seg 1/1 to: AMYA_METJA from: 1 to: 467

40 GAATCCCAAGAGCGAAATCCCAAGGTCATAGAGAAGGCATACATCC 89

```

34 LysLeuAsnLysGluValPheAsnLysValAlaAsnLysCysTyrIlePr 50
   ::::: ||||| :::: ||||| ::::: ||||| ::::: ||||| |||
90 AGTCATCGACACACTGATTA...GAAGAAATTCCTTTGGGCTCA 133
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
50 ofhrAsnGluLeuLeuGluLeuIleAspGluTyrAspPheLysValA 67
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
134 AC.....ATAACGGG...TATACCTTAAAGTTCCTCCCG 165
   || || || || || || || || || || || || || || ||
67 SerIleThrGlyValPheValGluGluInAlaLeuGluPhe..... 81
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
166 AAGGAT...ATTATAGACTCTGTTAAAGGGGGCATCGGAGTGACCTGAT 212
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
82 AsnAspTyrValLeuAspPheLysAspLeuValLysThrGlyAsnVa 98
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
213 AGAGATAATCGAAGAGCTACAGCGCAATACTCCCTCCCTCCCGC 262
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
98 lglLeuIleAlaGluThrTyrHisHisSerLeuThrSerLeuPheGluT 115
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
263 TT...AGCAGAGTAGAACACAACTTCAGAGATAGGGAAGTTAAGGAA 309
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
115 hrGluAspIlePheIleGluAspIleGluMetHisArgLysMetTyrLys 131
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
310 GAGCTCTTCGAGCTTTCCTCCAAAGGATTCCTGGCTGCCAGAGCTCGCTA 359
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
132 GluIlePheGlyPheLysAlaLysValPheArgAsnThrGluLeuIleTy 148
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
360 TGACCCGATATCCTCGCATCTACTGAAGGACACGGTTATGAGTATCAT 409
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
148 rasnAsnArgIleAlaLysIleAlaLysAspLeuGlyPheLysAlaIleP 165
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
410 TCGCGACGGG...GAGCGGATGCTTTCTCAGCTCATCTCAACTCGCG 456
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
165 heThrGluGlyIleGluLysIleLeu..... 173
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
457 ATAAAGCCAATTAACCGCTCTATCCACACTTATAAGGCCCAAGGGA 506
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
174 .....GlyIleArgSe 177
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
507 AAAGCGCTTTAGGTATACATCAG.....TATCTCTGCTGCTCAGGG 547
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
177 rProAsnTyrLeuTyrGlnSerProAspGlyMetLysIleLeuLeuArgA 194
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
548 ACCTTAGAAGCGGATAAGCTCGTGTGTTGAAGTAAGTAACCTTAAG 597
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
194 snTyrArgLeuSerAspIleGlyPheArg.....Phe 205
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
598 GCAGTCAAGACATCGAAGCGGTACCC.....GT 626
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
206 SerAlaArgAspTrpAspGlnTyrProLeuThrAlaAspLysTyrAlaI 222
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
627 TTGGGTGGCGGTGAACAGCGCTGTATCTCGGCATCGGAAGCTTCCTC 676
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
222 enrLeuAlaSerThrProGlyGluValIleAsnIle.....T 235
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
677 TTATGAATCCTAAGAAGTGGGAGC.....TGATAGAGACAAAGGAC 720
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
235 yMetAspTyrGluThrPheGlyGluHisHisTrpLysGlu..... 248
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
721 AACATTCCTTATACGGCAGCATATA...GAGTTCAATGGCTATAG... 765
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
249 .....ThrGlyIlePheGluPheLeuArgTyrLeuPr 259
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
766 ....GACATTGCAGCTAC...AGATGAGTGTTCAGGGATTATAGAG 808
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
259 orleGluIleAlaLysHisGluHisGluLeuValValAsnValSerGluV 276
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
809 TTATAGACGAGCTCAACTCGGAAGTGTGCTTCCTCCCTCAGAGCTGAAGCAC 858
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||
276 alValAspArgLeuGluProArg..... 283
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
859 AGTGGGAAGGAGCTCTACTTACGG.....ACTTCAGTTGGGCA... 897
   ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||

```

```

284 .....GlyGluIleTyrValHisGluPheAlaThrIleSerTrpAlaAs 298
898 ..CCAGATAAGACGTTGAGGATATGAGAGAGGACGAGGAGGACGAGAC 946
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
298 pThrGluArgAspValSerAlaTrpLeu.....GlyAsn...LysM 311
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
947 TTAATATGCTCTCTACAAATATGAGGGCGGAACCTGC...CTTTAGCC 993
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
311 etGlnArgIleSerPheGluLysLeuLysAspIleGlyLysPheIleLys 327
   ::::: ||||| :::: ||||| ::::: ||||| ::::: |||||
994 GAGAACAGCGATGCAAGGGGATGGAGCCCTCCCTCGAGAGAGGCTGGA 1043
   ||||| ||||| :::: ||||| ::::: ||||| ::::: |||||
328 GluAsnSerAsnLysLeu.....LysLysLeuAs 337
1044 TGCCTTCGGCGGATATATACGATTGGAGG 1074
   : ||| ||||| :::: ||||| ::::: ||||| ::::: |||||
337 mlyPheAspGluIleTyrLysMetTyrLys 347
seq_name: SwissProt_40:AMVA_PYRFU
seq_documentation_block:
ID AMVA_PYRFU STANDARD; PRT; 648 AA.
AC P49067;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-amylase (EC 3.2.1.1).
GN AMVA OR PF0272
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94043280; PubMed=8226990;
RA Lademan K.A., Asada K., Uemori T., Mukai H., Taguchi Y., Kato I.,
   Anfinsen C.B.;
RT "Alpha-amylase from the hyperthermophilic archaeobacterium Pyrococcus
   furiosus. Cloning and sequencing of the gene and expression in
   Escherichia coli.";
RL J. Biol. Chem. 268:24402-24407(1993).
[2]
CHARACTERIZATION.
MEDLINE=94043279; PubMed=8226989;
RA Lademan K.A., Davis B.R., Kruttsch H.C., Lewis M.S., Griko Y.V.,
   Privalov P.L., Anfinsen C.B.;
RT "The purification and characterization of an extremely thermostable
   alpha-amylase from the hyperthermophilic archaeobacterium Pyrococcus
   furiosus.";
RL J. Biol. Chem. 268:24394-24401(1993).
CC -!- FUNCTION: DISPLAYS A BROAD RANGE OF SUBSTRATE SPECIFICITY, WITH
   THE CAPACITY TO HYDROLYZE CARBOHYDRATES AS SIMPLE AS MALTOSE.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
   linkages in oligosaccharides and polysaccharides.
CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.
CC -!- SUBUNIT: HOMODIMER.
CC -!- MISCELLANEOUS: THE ISOELECTRIC POINT IS 4.3. THE ENZYME DISPLAYS
   OPTIMAL ACTIVITY, WITH SUBSTANTIAL THERMAL STABILITY, AT 100
   DEGREES CELSIUS, WITH THE ONSET OF ACTIVITY AT APPROXIMATELY 40
   DEGREES CELSIUS.
CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; L22346; AAA72035.1;
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57; 1.
KW Hydrolase; Glycosidase; carbohydrate metabolism.

```

```

FT INIT MET 0 0
SQ SEQUENCE 648 AA; 76178 MW; BF7A495F084E0FB1 CRC64;

alignment_scores:
  Quality: 141.50 Length: 391
  Ratio: 0.773 Gaps: 23
  Percent Similarity: 46.803 Percent Identity: 22.762

alignment_block:
US-09-886-400-3 x AMYA_PYRFU ..
Align seg 1/1 to: AMYA_PYRFU from: 1 to: 648

73 GACAAGGCATACATCCGCTGCTGAGACACATGATTAAAGAGAATATCC 122
|||||:||||| ||| :|||||: |||
28 GluLysCysTyrTrpProPheLeuGluThrLeu.....GluLysTyrPr 42
123 T.....TTTGGGCTCAACATACAGGGCTATACCTTAAGTTCTCC 163
| :|||||: ||| :|||||: |||
42 oAsnMetLysValAlaIleHisThrSerGlyProLeuIleGluTrpLeuG 59
164 CG.....AAGGATATATAGACTCGTTAAAGGGGCGATCGGAGT 204
|||||:|||||: |||: |||: |||: |||: |||: |||: |||: |||
59 lnAspAsnArgProGluTyrIleAspLeuLeuArgSerLeuValLysArg 75
205 GACCTGATAGATATACGGAACAGAGCTACACGCGCAATACCTCCCT 254
|||||:|||||: |||: |||: |||: |||: |||: |||: |||: |||
76 GlyGlnValGluIleValAlaGlyPheTyrGluProValLeuAlaLase 92
255 CTTCCGCTTAGCAGATGACACCAAGCTTCAGAGATAGGGAAGTGA 304
|||||: ||| :|||||: |||: |||: |||: |||: |||: |||: |||
92 rIlePro.....LysGluAspArgIleGluGlnIleArgLeuMetL 106
305 AGGAA.....GAGCTCTTCGAGCTTCTCCAAAGGATTCGGCTGCCA 348
|||||: ||| :|||||: |||: |||: |||: |||: |||: |||: |||
106 ySglnTrpAlaLysSerIleGlyPheAspAlaArgGlyValTrpLeuThr 122
349 GAGTCGCTATGACCGGATATCCCTGCCATCTAGAGCAACGGTGA 398
||| :|||||: |||: |||: |||: |||: |||: |||: |||: |||
123 GluArgValTrpGlnProGluLeuValLysThrLeuLysGluSerGlyTr 139
399 TGAGTATCTATTGCGCGAGCGGGGCGGATGCTTTTTCAGCTCATCTCA 448
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
139 eAspTyrValIleValAsp.....AspTyrHisPheM 150
449 ACTCGCGGATAAAGCCCAATTAACCGCTCTAT...CCACACCTTATAAG 495
|||||: ||| :|||||: |||: |||: |||: |||: |||: |||: |||
150 etSerAlaGlyLeuSerLysGluGluLeuTyrTrpProTyrTrpThrGlu 166
496 GCCCAAGGAA.....AAGCGCTTAGGTA 521
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
167 AspGlyGluValIleAlaValPheProIleAspGluLysLeuArg.. 182
522 CATCAGTATCTCTTGCTGCTCAGGAGCTTAGAAGCGGATAAAGCTC. 570
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
183 .....TyrLeuIleProPheArgProValAspLysValLeuGluTyrL 197
571 .....GTTTGTAGCTAAGCTAACGTAAGCGAGTC.....AAA 606
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
197 euHisSerLeuIleAspGlyAspGluSerLysValAlaValPheHisAsp 213
607 GACATCAAGCCGTACCGCTTTGGGTGGCGGTGAACACGCGCTGTAATGCT 656
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
214 AspGlyGluLysPheGlyIleTrpProGlyThrTyrGluTrpValTyr.. 229
657 CGGATCGGAGGCTTCTCTTATGAATCTTAAGAAAGTGGCGAGTGA 706
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
230 .....GluLysGlyTrpL 234
707 TA.....GAGGACAAAGGACACATCTCT 729
||| :|||||: |||: |||: |||: |||: |||: |||: |||: |||
234 euArgGluPhePheAspArgIleSerSerAspGluLysIleAsnLeuMet 250

```

```

730 CTATACGGCACCGCATATAGAG.....TTCATTGGCTATAG 764
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
251 LeuTyrThrGluTyrLeuGluLysTyrLysProArgGlyLeuValTyrLe 267
765 GGACATTGCGGCTACAGAATGAGTGTTCGAGGATATTAGAGTTATAG 814
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
267 uProIleAlaSerTyr.....PheGluMet.... 275
815 ACGAGCTCAACTCGGAACCTGTCCTTCCC..... 843
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
276 .....SerGluTrpSerLeuProAlaLysGlnAlaArgLeuPhe 288
844 .....TCAGAGCTGAAGCACAGTGA.....AGGGA 869
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||
289 ValGluPheValAsnGluLeuLysValLysGlyIlePheGluLysTyrAr 305
870 GCTCTACTTACGAGCTTCGAGTTGGCCACCATAGAGCTTGAGGATAT 919
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
305 gValPheValArgGlyGlyIleTrp.....LysAsnPhe...PheT 318
920 GGAGAGGAGGACGAAGGGAAC.....GCAAGACTTAATATGCTGTCC 960
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
318 yrlLysTyrProGluSerAsnTyrMetHisLysArgMetLeuMetValSer 334
961 TACAATATGAGGGCGGAACCTCCCTTTTAGCCGAGAACAGCGATGCAAG 1010
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
335 LysLeuValArgAsnAsn..... 340
1011 GGGATGGAGCCCTCCCTGAGAGGAGGCTGATGCTTCGGGGCGATAT 1060
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
341 .....ProGluAlaArgLysTyrLeuLeuArgAlaGlnC 352
1061 ATAACGAT.....TCGAGGGGT 1077
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||
352 ysAsnAspAlaTyrTrpHisGly 359
seq_name: SwissProt_40:MALQ_PYRKO
seq_documentation_block:
ID MALQ_PYRKO STANDARD; PRT: 653 AA.
AC O32450;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase)
DE (Disproportionating enzyme) (D-enzyme).
OS Pyrococcus kodakaraensis.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX NCBI_TaxID=69014;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOD1;
RA Tachibana Y., Fujiwara S., Takagi M., Imanaka T.;
RT "Cloning and expression of the 4-alpha-glucanotransferase gene from the hyperthermophilic archaeon Pyrococcus sp. KOD1, and characterization of the enzyme."
RT J. Ferment. Bioeng. 83:540-548(1997).
RL -!- CATALYTIC ACTIVITY: Transfers a segment of a (1,4)-alpha-D-glucan to a new 4-position in an acceptor, which may be glucose or (1,4)-alpha-D-glucan.
CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D87907; BAA22062.1;
CC InterPro; IPR004300; Glyco_hydro_57.
DR

```

DR Pfam; PF03065; Glyco_hydro_57; 1.
KW Transferase; Glycosyltransferase; Carbohydrate metabolism.
SQ SEQUENCE 653 AA; 76664 MW; B0C4695613F29219 CRC64;

alignment_scores:
Quality: 121.50 Length: 386
Ratio: 0.653 Gaps: 21
Percent Similarity: 48.187 Percent Identity: 21.244

alignment_block:

US-09-886-400-3 x MALQ_PYRKO ..

Align seg 1/1 to: MALQ_PYRKO from: 1 to: 653

```
73 GAGAGGATACATCCAGTCTATCGACACATGATTAAGAGAAATTC 122
|||||:||||| ||| :|||:||||| ||||| ||
28 GluArgSerTyrArgProPheMetGluThrLeu.....GluGluTyrPr 42
123 T.....TTGGGCTCAACATAACGGGCTATACCTTAAAGTTCCTC. 162
|:|||||:||||| ||||| |||||:|||||
42 OasnMetLysValAlaValHisTyrSerGlyProLeuLeuGluTrpIleA 59
163 .....CCGAGGATATTATAGACCTCGTTAAAGGGGCGATCGCGAGT 204
||:|||||:|||||:|||||:|||||:|||||
59 rGAspAsnLysProGluHisLeuAspLeuLeuArgSerLeuValLysArg 75
205 GACCTGATAGAGATAATCGGACGAGCTACACGACGCAATACTCCOCT 254
|||||:||||| |||||:||||| |||||:|||||
76 GlyGlnLeuGluIleValAlaGlyPheTyrGluProValLeuAlaIse 92
255 CTTCCCGTTAGCAGAGTAGAGACACAGTTTCAGAGAGATAGGCAAGTTA 304
|||||:||||| ||| :|||:||||| ||||| |||||
92 rIleProLysGluAspArgIleValGlnIleGluLysLeuLysGluPheA 109
305 AGGAAGAGCTCTCGAGCTTCTCCAAAGGATTCTGGCTCCAGAGCTC 354
|||||:||||| |||||:||||| |||||:|||||
109 laArgAsnLeu...GlyTyrGluAlaArgGlyValTrpLeuThrGluArg 124
355 GCTATGACCGGATATCCCTGCCATCTAGGACACACGGTTATGAGTA 404
|||||:||||| ||| :|||:||||| ||||| |||||
125 ValTrpGlnProGluLeuValLysSerLeuArgAlaAlaGlyIleAspTy 141
405 TCTATTCCGCGAGCGGAGCGATCTTTCTCAGCTCATCTCAACTCGG 454
||:|||||:||||| ||| :|||:||||| ||||| |||||
141 rValIleValAsp.....AspTyrHisPheMetSerA 152
455 CG...ATAAGCCAATTAAACGGCTCTATCCACACCTTATAAGGCCAA 501
||:|||||:||||| ||| :|||:||||| ||||| |||||
152 laGlyLeuSerLysAspGluLeuPheTyrProTyrTyrThrGluAspGly 168
502 AGGGA.....AAGCGCTTAGGTACATCAG 527
||| |||||:||||| |||||:|||||
169 GlyGluValIleThrValPheProIleAspGluLysLeuArg..... 182
528 CTATCTCTCTGCTCAGGAGCTTAGGAAGCGGATAAGCTCGTTT. 576
|||||:||||| ||| :|||:||||| ||||| |||||
183 TyrLeuIleProPheArgProValAspLysThrLeuGluTyrLeuHis 199
577 .....GAAGTAGGTAAAGCTTAAGGCGCTC.....AAGACATC 612
|||||:||||| ||| :|||:||||| ||||| |||||
199 erLeuAspAspGlyAspGluSerLysValAlaValPheHisAspAspGly 215
613 GAAGCGGCTACCGCTTGGGTGGCGCTGAACAGCGCTGTAATGCTCGGAT 662
||||| |||||:||||| |||||:|||||
216 GluLysPheGlyValTrpProGlyThrTyrGluTrpValTyr..... 229
663 CGGAAGGCTTCTCTTATGATCTTAAGAACTGGCGAGCTGGATA.... 708
|||||:||||| ||| :|||:||||| ||||| |||||
230 .....GluLysGlyTrpLeuArg 236
709 .....GAGGACAAGGACAACATCTTCTATAC 735
|||||:||||| |||||:||||| |||||:|||||
```

```
236 luPhePheAspArgValSerSerAspGluArgIleAsnLeuMetLeuTyr 252
736 GGCACCGCATATAGAG.....TTCAATTGGCTATAGGACAT 770
|||:|||||:||||| ||| :|||:||||| ||||| |||||
253 SerGluTyrLeuGlnArgPheArgProArgGlyLeuValTyrLeuProIle 269
771 TGCAGGCTAC...AGAATGAGTGTGAGGATTA..... 801
|||||:||||| |||||:||||| |||||:|||||
269 eAlaSerTyrPheGluMetSerGluTrpSerLeuProAlaArgGlnAlaL 286
802 .....TTAGAGGTTATAGCAGCTCAACTCGAACTGTGCCTTCCC 843
|||:|||||:||||| |||||:||||| |||||:|||||
286 ysLeuPheValGluPheValGluLeuLys..... 296
844 TCAGAGCTGAAGCAGTGGAGGAGCTCTACTAGGACTTCGAGTTG 893
|||||:||||| |||||:||||| |||||:|||||
297 LysGluAsnLysPheAspArgTyrArgValPheValArgGlyIleTyr 313
894 GGCACAGATAAGAGCTTGAGGATATGGAGAGGACGAGGGAAC.... 939
|:|||||:||||| |||||:||||| |||||:|||||
313 p.....LysAsnPhe...PhePheLysTyrProGluSerAsnTyrM 326
940 .....GCAGACTTAATATGCTGCTACATATAGGAGGCGGACTCGCC 984
|||||:||||| |||||:||||| |||||:|||||
326 eHisLysArgMetLeuMetValSerLysAlaValArgAsnAsn..... 340
985 CTTTTCGCGAAGCAGCGATGCAAGGGGATGGAGCCCTCCCTGAGAG 1034
|||||:||||| |||||:||||| |||||:|||||
341 .....ProGluAl 343
1035 GAGGTGATGCCCTCCGGCGATATATAAGAT.....TGGAGGGGTG 1078
|||||:||||| |||||:||||| |||||:|||||
343 argGluPheIleLeuArgAlaGlnCysAsnAspAlaTyrTrpHisGlyV 360
1079 AAAATGGG 1086
|||
360 alPheGly 362

seq_name: SwissProt_40:MALQ_THELI
seq_documentation_block:
ID MALQ_THELI STANDARD; PERT; 659 AA.
AC O32462;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase)
DE (Disproportionating enzyme) (D-enzyme).
GN JGT.
OS Thermococcus litoralis.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX NCBI_TaxID=2365;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 1-19 AND 427-437, AND
RP CHARACTERIZATION.
RC STRAIN=DSM 5473;
RA MEDLINE=97454300; PubMed=9310375;
RX Jeon B.-S., Taguchi H., Sakai H., Ohshima T., Matsuzawa H.;
RT "4-alpha-glucanotransferase from the hyperthermophilic archaeon
RT Thermococcus litoralis. Enzyme purification and characterization, and
RT gene cloning, sequencing and expression in Escherichia coli.";
RL Eur. J. Biochem. 248:171-178(1997).
CC -!- FUNCTION: Catalyzes the transglycosylation of
CC maltooligosaccharides, yielding maltooligosaccharides of various
CC lengths and glucose.
CC -!- CATALYTIC ACTIVITY: Transfers a segment of a (1,4)-alpha-D-glucan
CC to a new 4-position in an acceptor, which may be glucose or (1,4)-
CC alpha-D-glucan.
CC -!- ENZYME REGULATION: INHIBITED BY P-CHLOROMERCURIBENZOIC ACID,
CC MONIODOACETIC ACID, MERCURY AND NICKEL IONS.
CC -!- MISCELLANEOUS: OPTIMAL ACTIVITY IS FOUND AT 90 DEGREES CELSIUS.
CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
-----
```

610 ATCGAAGCCGTACCCGTTTGGGTGGCCGTGAACACGCTCTAAATGCTCGG 659
 :: ||| ||| :: :: :: ::
 203 AspAspProSerLysValAlaValPheHisAspAspGlygluLysPheGl 219
 :: :: :: :: :: :: ::
 560 CATC.....GGAAGGCTTCCTCTTAATCTCAAGAAAGTGCCAGACT 703
 :: ||| :: :: ||| :: ::

219 yValTppProGlyThrTyrgluTpValTyrgluLys.....Glyt 238

704 GGATA.....GAGCACAGGCACANTT 726
||||| :||| ||||| |

233 rIpeuArgluPheAspAlaIleThrSerAsnGluyLiIeAsnLeu 249

727 CTCTATACGCCACCGCATATAGACTTCATTGGC..... 759

MetInrYrSer.....GluTyrLeuSerLysPheThrProArgGI 26
760TATAGGCACATTCCAGCTACAGAATCAGTCTTGAGGGATTAT 802

vLeuValTyrLeuProIleLaserTyr.....P 273

803 TAGAGGTTATAGACGAGGCTCAACTCGGAACCTGTGGCTTCCCTCA 846

803 TAGAGGIIATAGACGAGCICAACTCGGAACIGTGCCCTTCCCCICA...: 849

::::: ||||| :::::

273 NeglumeL.....SERGIUPSELEUPTOATALYSUM 208

847GAGCTGAAGCACAGTGGAGG... 86
285 AlaLysLeuPheValGluPheValGluClnLeuLysGluGluGlyLysPh 301
868GAGCTCTACTTACGCGAATTCGAGTGGGACCACGATAAGA 907
301 eGluLysTyrArgValPheValArgGlyGlyLeuTrp.....LysA 315

908 GCTTGGAGTATTGGAGAGGACCAAGGGAAC.....GCAAGACACTT 948
 315 snPhe...PhePheLysTyrProGluSerAsnPheMetHisLysArgMet 330
 949 AATATGCTCTCTCTACAAATATGAGGGGGGAACTCGCCCTTTTACCCGAGAA 998

331 LeuMetValSerLysAlaValArgAspAsn..... 348
999 CACCGATGCAAGGGATGGGAGCCCTCCCTGACGAGAGGCTGGATGCCT 104
341 348
ProGluAlaArgLysTyrIleLeu 348

name: SwissProt_40:YG1F_YEAST

```

Documentation_Block:
  YGIF_YEAST STANDARD; PRT; 551 AA.
  P53214;
  01-OCT-1996 (Rel. 34, Created)
  01-OCT-1996 (Rel. 34, Last sequence update)
  15-DEC-1998 (Rel. 37, Last annotation update)

```

YGR023W.
Saccharomyces cerevisiae (Baker's yeast).
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_taxID=4932;

SEQUENCE FROM N.A.
STRAIN=S288;
MEDLINE=97435481; PubMed=9290212;
Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
"Sequence analysis of 203 kilobases from *Saccharomyces cerevisiae* chromosome III." *Genomics* 47:1-10 (1997).

Yeast 13:1077-1090(1997).
 -!- SIMILARITY: TO YEAST MID2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: Z72807; CAA97006.1; -
DR SGD: S0003255; MTL1.
KW Hypothetical protein; Transmembrane.
FT DOMAIN 104 107 POLY-SER.
FT DOMAIN 111 120 POLY-SER.
FT DOMAIN 124 138 POLY-SER.
FT DOMAIN 142 146 POLY-SER.
FT DOMAIN 182 188 POLY-SER.
FT DOMAIN 215 218 POLY-SER.
FT DOMAIN 222 232 POLY-SER.
FT DOMAIN 236 264 POLY-SER.
FT DOMAIN 269 275 POLY-SER.
FT DOMAIN 278 281 POLY-SER.
FT DOMAIN 284 291 POLY-SER.
FT TRANSMEM 362 382 POTENTIAL.
FT DOMAIN 468 474 POLY-ASP.
SQ SEQUENCE 551 AA; 57527 MW; 86D94341B319951 CRC64;

alignment_scores:

Quality: 115.50 Length: 290
Ratio: 0.791 Gaps: 10
Percent Similarity: 50.345 Percent Identity: 25.172

alignment_block:

US-09-886-400-3/rev x YGIF_YEAST ..

Align seg 1/1 to: YGIF_YEAST from: 1 to: 551

995 TCGGCTAAAGGCGGCTGCGCCCTCATATTGTAGGACAGCATATTAAAG 946
||||| ||||| ||||| |||||
69 SerSerAlaSerLeuSerThrProSerIleAla.....SerValSerPh 83
945 TCTTGGTTCCCTTCGTCTCTCTCCATCTCAAGCTCTTATCTGTGT 896
||||| ||||| ||||| ||||| |||||
83 eThrSerPheProGlnSerSer...SerLeuLeuThrLeuSerThrL 99
895 CCCAACTCGAAGTCGTAGTAGAGCTCCCTCCACTGTGCTTCAGCTCT 846
||| ||||| ||||| ||||| ||||| |||||
99 eUerSerGluLeuSerSer...SerSerMetGlnValSerSerSer 114
845 GAGGGAAGCAGCTCCGAGTGT.....AGCTGCTCTAT 811
||| ||||| ||||| ||||| ||||| |||||
115 ThrSerSerSerSerSerGluValThrSerSerSerSerSerSerI 131
810 AACCTCTATATCCCTCAACACTCATCTGTAGCTGCAATGCTCCCTAT 761
||||| ||||| ||||| ||||| ||||| |||||
131 eSerProSerSerSerSerSerThrIle..... 141
760 AGCCAATGAACCTATATCGGTGCGGTATAGAAGATGTTGCTGTGCC 711
||| ||||| ||||| ||||| ||||| |||||
142SerSerSerSerLeuProThrPheThrValAlaSerThrSer 156
710 TCTATCCAGCTGCCACTTCTTAGATCATATAGAGAACCTTCCGAT 661
||| ||||| ||||| ||||| ||||| |||||
157 SerThrValAlaSerSerThrLeuSer...ThrSerSerSerLeuVal 172
660 GCGGACATTACAGCGGTGTCACGGCCACCAACAGGTCAGCGCTTCA 611
||| ||||| ||||| ||||| ||||| ||||| |||||
172 eSerThrSerSerSerThrPheSerSerGluSerSerSerSerL 189
610 TGTCTTGAAGCTTACGTTACCTTACCTTCAAAACGAGCTTTATC 561
||| ||||| ||||| ||||| ||||| ||||| |||||
189 eUleSerSerIleSerThrSerValSerThrSerSerValThrVal 205

560 GCCTTCCTAAGCTCCCTGAGACCAAGAGATAGCTGATGTACCTAAAGCG 511
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
206 ProSerSerSerThrSerSerProPro..... 214
510 CTTTCCCTTTGGGCTTTATAGGTGTGGATAGAGCGGTTTAAATGGCT 461
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
215SerS 216
460 TTATCGCGAGTTGAGATGAGTGAAGAAAGCATCGCTCCCGCTCGGCG 411
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
216 eSerSerGluLeuThrSerSerSerTyrSerSerSerSerSerSerSer 232
410 AATAGATACTATACACCGTTGCTCTTCAGTATGGCAGGATATATCGGCTC 361
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 ThrLeuPheSerTyrSerSerPheSer..... 242
360 ATAGCGGAGCTGCGCAGCCAGATCCCTTTGGAGAAAGCTCGAAGAGCT 311
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
243SerSerSerSerSerSerSerSerSerSerSerSerSerS 258
310 CTTCTTAACTTCCCTATCTCTGAACTTGTGCTTACTCTCTCTAAGC 261
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
258 eSerSerSerSerSerTyr.....PheThrLeuSerThr 270
260 GAGAGAGGGGAGTATTCGTCGTGTAGCTCGTTCGGATATCTCTAT 211
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
271 SerSerSerSerSerIleTyrSerSerSerTyrProSerPheSerSe 287
210 CAGGTCACCTCGGATCGCCCTTTAACGAGGTCTATATATATCTCTCGGA 161
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 rSerSerSerSerAsnPro.....ThrSerSerIleThrSerThrSera 302
160 GGAACCTTAAAGTATAGCC 141
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
302 laSerSerSerIleThrPro 308
seq_name: SwissProt.40:MDH_BAGSU
seq_documentation_block:
ID MDH_BAGSU STANDARD; PRT; 311 AA.
AC P49814;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37) (Vegetative protein '69) (VEG69).
GN CITH OR MDH.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / SMV;
RX MEDLINE=95134995; PubMed=8550482;
RA Jin S., de Jesus-Berrios M., Sonenshein A.L.;
RT "A Bacillus subtilis malate dehydrogenase gene."
RL J. Bacteriol. 178:560-563(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=98048467; PubMed=9387221;
RX Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequencing and functional annotation of the Bacillus subtilis genes
in the 200 kb rnaB-dnaB region."
RL Microbiology 143:3431-3441(1997).
RN [3]
RP SEQUENCE OF 1-15.
RC STRAIN=IS58;
RX MEDLINE=97443988; PubMed=9298659;
RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
RA Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
regulation map for Bacillus subtilis."
RL Electrophoresis 18:1451-1463(1997).


```

30 laAlaSerThrSerLeuSerSerSerVal..... 40
686 GGATTCAAGAGGAGCGTCCGATCCGAGCATACAGCCGTGTCAC 637
41 .....ilProSerSerSerSerMetLeuSerSe 51
636 GCCACCAACAGGGTAGCGTTCGATGTC.....TTGACTG 599
..... 511 ..... 51
51 rSerSerAlaThrAlaLeuSerSerSerSerSerProLeuSerS 68
598 CTTTACGTTACCTTACCTCAAAAGGAGCTTATCGCCTTCCTAAGC 549
.. 511 ..... 511 ..... 511
68 exSerSerPheThrSerProAlaSerSerPheLeuThrSerLeuVal 84
548 TCCCTGAGACCAAGAGATAGCTGATGATACCTAAAGGCGCTTTCCTTGG 499
511 ..... 511 ..... 511
85 SerSerSerSerGln.....GlnSerSerSerSe 95
498 GCCCTTTATAGGTGATAGAGCGGTTTAATTTGGCTTATCCCGAGT 449
511 ..... 511 ..... 511
95 rAlaSerLeuThrSerSerSerAlaThrLeu..... 106
448 TGAGATGAGCTCAGAAAGCATCGCTCCCGTCGGCGAATAGATACTCA 399
511 ..... 511 ..... 511
107 .....ThrSerSerSerAlaSerProThrSerSerSerSer 120
398 TAACCGTTCTCCTCAGTATGCGAGGATATCGGTCATAGCGGAGCTC 349
511 ..... 511 ..... 511
121 HisAlaLeuSerSerSerSerSerLeuValAlaSerSerSerSe 137
348 TGGC.....AGCCAGAATCCCTTTGGAG 326
511 ..... 511 ..... 511
137 rGlyMetSerSerSerSerLeuSerHisSerSerSerValProSerSers 154
325 AAAGCTG...AAGAGCTCTTCTTAACCTCCCATCTCTCTGAACTGT 279
514 exSerSerTyrHisSerSerSerSerSerSerSerSerSerSerSer 170
278 GCTTCTACTCTGCTAAGCGGAGGAGGGAGTATCGCTCGGTAGCT 229
511 ..... 511 ..... 511
171 AlaSerIleValSerSerThrTyrArgAspGlyProSerIleThrLe 187
228 CGTTCGGATTATCTATCAGGTCTCCTCGCATCGCTCCCTTTAACG... 183
511 ..... 511 ..... 511
187 uValSerThrSerTyrValSerGluValValThrProThrThrAsnA 204
182 .....AGGTCTATAATCTCTCGGGAGAACTTTAAGTATAGCCGCTT 138
204 snTrpAsnSerSerSerSerPheThrSerSerThrSerThrProile 220
137 ATGTTGAGCCCAAGAAATTTCTTTTAACTACAGTCTCGATGACTGG 88
221 SerSerSerTyrSerSerSerGlyThrLeuProSerSerSerSerAsnLysSe 237
511 ..... 511 ..... 511
87 GATGATCGCTTCTCTATGACCTTTGGATTTCG..... 54
237 rSerAsnHisValGlyValValValGlyCysSerValAlaIleProValG 254
53 .....CTCTTTGGATTTCGGCATAC...TGGAGG 27
254 llyValValLeuLeuLeuLeuGlyLeuGlyIlePheLeuTrpLys 268
seq_name: SwissProt_40:MDH_BACHD
seq_documentation_block:
ID MDH_BACHD STANDARD; PRT; 314 AA.
AC Q9K849;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
GN MDH OR CITH OR BH3158.

```

```

OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) -> oxaloacetate + NADH.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP001517; BAB06877.1; -
DR InterPro: IPR001357; L_LDH.
DR InterPro: IPR001252; MDH_actsite.
DR InterPro: IPR000205; NAD_binding.
DR InterPro: IPR001236; ldh.
DR Pfam: PF00056; ldh; 1.
DR Pfam: PF02866; ldh.C; 1.
DR PRINTS: PR00086; LLDHGRNASE.
DR PROSITE: PS00086; MDH; FALSE_NEG.
DR Oxidoreductase; Tricarboxylic acid cycle; NAD; Complete proteome.
KW ACT_SITE 153 153 PROTON-RELAY (BY SIMILARITY).
FT BINDING 156 156 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).
FT ACT_SITE 180 180 PROTON-RELAY (BY SIMILARITY).
SQ SEQUENCE 314 AA; 33680 MW; 02D132F11B3B8E34 CRC64;

alignment_scores:
  Quality: 103.50      Length: 217
  Ratio: 1.005        Gaps: 9
Percent Similarity: 47.465 Percent Identity: 23.041

alignment_block:
US-09-886-400-3 x MDH_BACHD ..
Align seg 1/1 to: MDH_BACHD from: 1 to: 314

28 CTCACGATGCGGAATCCCAAGAGGGAATCCCAAG..... 66
130 MetThrTyrThrValTyrLysGluSerGlyPheProLysAsnArgValI 146
67 .....GTCAATAGAGAGGCATACATCCAGTCATCGAGACAC 103
146 eclyGlnSerGlyValLeuAspThrAlaArg.....PheArgThrP 160
104 TCATTAAAGAGAATTCCTTTGGGCTC...AACATAACGGGCTATACC 150
160 heValAlaGlnGluLeuAsnLeuSerValGluAspIleThrGlyPheVal 176
151 TTAAGATTCCTCCCGAGGATATTATTAGAACCTCGTTAAAGGGGCGCATCG 200
177 Leu.....GlyGlyHisG 181
201 GAGTCACCTGATAGAGATAATCGGAACGAGCTACACGACGACATA.... 246
181 yAspAspMetValProLeuIleArgTyrSerTyrAlaGlyGlyIleProL 198
247 ..CTCCCGCTCCTCCCGCTTACGAGAGTAGAAGCACAAAGTTCAGAGAGAT 294

```

```

198 euGluLysLeuLeuProGlnGluArgIleAspAlaIleValGluArgThr 214
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
295 AGGAAGTTAAGGAAGAGCTCTCGAGCTTTCCAAAGGA.....TT 338
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
215 ArgLysGlyGlyGlyIleValGlyLeuGlyAsnGlySerAlaIle 231
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
339 CTGGCTGCCAGAGCTCGCTATGACCCGATAATCCCTGCCATCTGAAGG 388
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
231 rTyAlaProAlaAlaSerLeuAlaGluMetValGluAlaIleLeuLysA 248
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
389 ACAACGGTTATGAGTATCTATCTCGCGACGGGAGCGATGCTTTCTCA 438
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
248 sPlys..... 249
439 GCTCATCTCACTCGCGGATAAGCAATTAACCGCTCTATCCACACT 488
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
250 .....LysArgValLeuProThrIleAlaTyrLe 259
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
489 TATAAAGGCCCAAGGAAGCGCTTAGGTAC..... 522
259 u.....GluGlyGlyTyrGlyLeuAspIleTyrValG 271
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
523 .ATCAGCTATCTCTGTCTCAGGAGCTTAGGAAGCGGATAAGCTC 570
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
271 lyValProThrIleLeuGlyGlyAspGlyIleGlyLysValIleGluLeu 287
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
571 GTTTTCAGGTAGTACGCTAAGGAGCAGTCAAGACATGGAAGCGGT 620
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
288 AspLeuThrAspGluGluLysAlaThrPheAlaLysSerIleGluSerVa 304
621 A 621
304 1 304

```

seq_name: SwissProt_40:YNR6_YEAST

seq_documentation_block:

```

ID YNR6_YEAST STANDARD; PRT; 636 AA.
AC P53882;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 67.4 kDa protein in RPS3-PSD1 intergenic region.
GN YNL176C OR N1661.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Obermaier B., Piravandi E., Rinke M., Dondey H.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 271452; CAA96068.1; -
DR SGD; S0005120; YNL176C.
KW Hypothetical protein; Transmembrane.
FT DOMAIN 94 98 POLY-SER.
FT DOMAIN 131 135 POLY-SER.
FT DOMAIN 301 309 POLY-SER.
FT DOMAIN 317 322 POLY-SER.
FT TRANSMEM 457 477 POTENTIAL.
SQ SEQUENCE 636 AA; 67356 MW; 4027EF48DEA5F2B7 CRC64;

```

```

alignment_scores:
  Quality: 103.50      Length: 343
  Ratio: 0.681        Gaps: 16
  Percent Similarity: 44.315  Percent Identity: 23.324

alignment_block:
US-09-886-400-3/rev x YNR6_YEAST ..
Align seg 1/1 to: YNR6_YEAST from: 1 to: 636

1088 TCCCCATTTTCCACCCCTCCCAATCGTTATATATATCGCCGGAAGGCATCCAG 1039
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
105 SerGlnTyrAspProAlaThrSerSerTyr.....Se 115

1038 CTCCTCTCTCAGGAGGGCTCCCATCCCTTCGATCGCTGTTCCTCGGCTA 989
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
115 rIleIleThr.....ProSerMetSerIlePheSer.... 125

988 AAAGGCGGAGTTCCGCCCTCATATTGTAGGACAGCATATTAGTCTTCG 939
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
126 .....SerThrSerProMet.....SerSerSerSerIleThr 137

938 TTCCTCTCTCTCTCTCATATTCCTCAAGCTCTTATCTGTGGTGCCTCACT 889
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
138 SerGluTyrSerSerLeuThr..... 144

888 CGAAGTCCGTAAGTAGAGCTCCCTCCACTGTCTCAGCTCTGAGGAA 839
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
145 .....SerThrProThrLeuSerSerSer..... 153

838 GGCACAGTTCGAGTGTAGCTCGTCTATAACCTCT...AATAATCCCTCA 792
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
154 .....AlaThrSerLeuSerSerSerTyrSerSerLeuSerSerProSer 168

791 AACTCATTTCTGTAGCTGCAATGTCCTATAGCCCAATGACTCTATATC 742
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
169 SerLeuLeuValSerSerSerLeuSerLeu.....SerLeuSe 181

741 GGTCCGCTATAGAAGAATGTTCTCTCTATCCAGCTCGCCACTT 692
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
181 rSerSerTyr.....SerAspThrLysLeuPheSerP 192

691 TCTTAGAATTCATAGGAAGCTTCGATCGGAGCATTTACAGCCGTG 642
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
192 heAspSer.....ArgSerSerIlePheSerProSerThrProThrVal 206

641 TTCAGGCCACCAAGCGGTACGCTTCGATGTCTTTGACGCTTTAG 592
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
207 IleSerProSerTyrThrTyrLeuSerSerIleSerAlaThrSerPheG 223

591 CGTTACTTACCTTCAAAACGAGCTTTATCGCCTTCCTTAAGCTCCCTGA 542
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
223 nIleSerThrThrSerGluLeuSerSerSerTrpPheSerThrIleSerS 240

541 GACCAAGGAGATAGCTGATGATACCTAAAGCGCTTTTCCCTTTGGGCTTT 492
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
240 erProSer..... 242

491 ATAAGTGTGGATAGAGCGGTATTAAATGGCTTTATCGCCGAGTTGAGATG 442
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
243 .....ThrIl 244

441 AGCTGAGAAAAGCATCGCTCCCGCTCGCGGAATAGATACTCATACCGT 392
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
244 eSerAsnLysAspThrThrCysProSerSerSerArgAsnThrSerThrS 261

391 TGTCTCTTACATATGGCAGGATATATCGGCTCATAGCGAGCTCTGGCAGC 342
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
261 erPheTyrSer.....SerSerLeuSerSer 269

341 CAGATATCCCTTTGGAGAACGCTCGAAGAGCTCTTCTCTTAACCTCCCTATC 292
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
270 ThrAsnAspPheSerThrIleSerLysSerSerLysLeuSerProSerAl 286

```

334ThrValGluMetVa 338

```
573 TTTTGAAGTAAAGTAAAGCTAAAGCAGCTAAAGACATCGAAGCGTA. 621
||||| : : : |||||
338 lPheAspGluSerGlyAsnLeuLysAlaTyrAspIleTyrGluSerValI 355
622 .....CCGGTTGGTGGCGTGAACACGGCTCTAATGCTCGCATC 663
||||| : : : |||||
355 leArgSerLysAlaArgLeuThrTyrAsnGluAlaLeuAlaLeuVal 371
664 GGA.....AGCCTCTCT...CTTATGAATCTTAAGAA 692
||||| : : : |||||
372 GlyAspProAlaLeuGluLysLysPheProAsnLeuValGluProLeuAr 388
693 AGTGGCGGAGC.....TGG..... 705
388 gMetMetGluThrLeuTyrArgIleLeuSerArgLysArgTyrGluMetG 405
705 ..... 705
405 lySerIleAspPheAspLeuProGluAlaGluValIleValAspGluTyr 421
705 ..... 705
422 GlyGluProThrAlaIleTyrProTyrGluArgHisValAlaHisArgI 438
706 .....ATAGGACAGGACACATCTTCTATAGGCA 739
438 eIleGluHisPheMetIleSerAlaAsnGluThrValAlaLeuHis.... 453
740 CCGATATAGATTCATGCTATAGGACATTCAGGCTACAGATG.... 786
454 .....LeuGluHisAlaGlyTyrProCysLeu.....TyrArgValHis 466
787 .....AGTGTGAGGATTTATAGAGTTATAGACGA 818
467 GluProProAspGluGluLysValGluAsnLeuLeuGluIleLeuGlu 483
819 GCTCACTCGGACTGTCCTTCCCTCAGAGCTGAAGCACAGTGGAGGG 868
||||| : : : |||||
483 yLeuGlyTyrLysValLysArgProHisGlu..... 493
869 AGCTCTACTTACGGACTTCGAGTGGCCACAGATAAGAGCTTGAGGATA 918
494 .....TyrThrProLysPheGlnLysIle 502
919 TGGAGAGAGACGAGGAGGACGAACGACTTAATATGCTCTCTAC 963
||||| : : : |||||
503 lIleGluAspPheGluGlyArgProGluGluAsnLeuValArgPhe 517

seq_name: SwissProt_40: LONM_SCHPO
seq_documentation_block:
ID LONM_SCHPO STANDARD; PRT: 1067 AA.
AC Q09769;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative ATP-dependent protease, mitochondrial precursor
DE (EC 3.4.21.-).
GN SPAC22F3.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lye G., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR INTRAMITOCHONDRIAL PROTEOLYSIS. CATALYZES
CC THE INITIAL STEPS OF PROTEIN DEGRADATION (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (Potential).
```

```
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16; ALSO KNOWN AS THE
CC LON FAMILY OF ATP-DEPENDENT PROTEASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL: Z54285; CAA91071.1; ..
DR MEROPS: S16.002; ..
DR InterPro: IPR003593; AAA.
DR InterPro: IPR003599; AAA_subfam.
DR InterPro: IPR003111; LON.
DR InterPro: IPR001984; Lon_endopep.
DR Pfam: PF00004; AAA; 1.
DR Pfam: PF02190; LON; 1.
DR PRINTS: PR00830; ENDOLAPTASE.
DR SMART: SM00382; AAA; 1.
DR SMART: SM00464; LON; 1.
DR PROSITE: PS01046; LON_SER; 1.
KW Hypothetical protein; Hydrolase; Serine protease; ATP-binding;
KW Mitochondrion; Transit peptide.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN 1 1067 585 ATP (POTENTIAL).
FT NP_BIND 578 585 ATP (POTENTIAL).
FT ACT_SITE 946 946 BY SIMILARITY.
FT SEQUENCE 1067 AA; 118641 MW; 215FCFBE9C0C4AD CRC64;
SQ
```

alignment_scores:

Quality:	101.00	Length:	454
Ratio:	0.561	Gaps:	22
Percent Similarity:	39.648	Percent Identity:	21.145

alignment_block:

US-09-886-400-3 x LONM_SCHPO ..

Align seg 1/1 to: LONM_SCHPO from: 1 to: 1067

```
53 AAAGTCATAGAGAGCATACATC.....CCAGTCATCG 97
614 LysGlyHisArg.ArgThrTyrIleGlyAlaMetProGlyLysIleValG 630
98 AGACACTG.....ATTAAAGAAATTCCTTTGGCTCAACATAACG 141
630 lAlaLeuLysLysValGlnThrGluAsnProLeuIleLeu..... 643
142 GCCTATACCTTAAGTCTCCGCAAGGATATATAGACCTCGTT..... 186
644 .....IleAspGluIleAspLysValGly 652
187 .....AAAGGGGCATCGAGTCACCTGATAGAGATAATCGGAACGA 229
652 sSerHisGlnGlyAspProAlaSerAlaLeuGluLeuLeuAspSerG 669
230 GGTACAGCAGCGCAATCTCCCTC.....CTCCGCTTAGCAGA 270
669 luGlnAsnSerAlaPheLeuAspTyrTyrMetAspIleProLeu..... 683
271 GTAGAACCAAGTTTCAGAGAGATAGGAGATTAGAGAGAGCTCTCGA 320
684 .....AspValSerValLeuPheVa 691
321 GCTTCTCCAAAGGATCTGCTGCTGCAGAGCTCGCTATGACCCGATAA 370
691 lCysThrAlaAsn.....ThrIleAspThrIleP 701
371 TCCTGCGCATCTG.....AAGGACACAGGTAT... 399
701 roProProLeuLeuAspArgMetGluValIleGluLeuSerGlyTyrVal 717
```

```

400 .....GAGTATCTATTGCGCGACGGGA 422
718 SerAlaGluLysValAsnIleAlaLysGlyTyrLeuIleProGlnAlaLys 734
423 GCGGATG.....CTTTCTCAGCTCATCTCAACTCGCGCATAAAGCAA 466
734 salaalacysGlyLeuLysAspAlaAsnValAsnIleSerAspAlaI 751
467 TTAACCCGCTCTATCCACACCTTATAAGGCCCAAGGAAAGCGCTTT 516
751 leLysGlyLeu.....Ile 755
517 AGGTACATCAGCTATCTCTTGGTCTCAGGAGCTTAGGAAGCGCATAAA 566
756 SerTyrTyrAlaHisGluSerGlyValArgAsnLeuLysLysSerIleG 772
567 GCTCGTTTGAAGTAAAGTAAGCGCTAAAGCGAGCTCAAGACATCGAAG 616
772 uLysIlePheArg.....LysThrSerPheSerIleValLysGluIleAsp 788
617 CGGTACCCGCTTGGTGGCGCGTGAACAG..... 645
788 sp.....GluLeuAsnSerLysGluLysSerThrGlyLys 799
646 .....GCTGTAATGCTCGG 659
800 SerGlyLysLysThrSerProGlnSerSerGluAspAlaAlaAsnLysG 816
660 CATCGGAAGGCTCTCTATGAATCTCAAGAAAGTGGCGAGCTGGATAG 709
816 uAlaSerSerValProLeuLysValProAspLysValAsnIleGluIleG 833
710 AGCAACAAGAC.....ACATT 726
833 luGluLysAspLeuThrLysTyrLeuGlyProIleTyrThrSerGln 849
727 CTTCTATACGGCACCGAT.....ATAGAGTTTCATGTGCTATAG 764
850 ArgLeuTyrAspThrThrProGlyValValMetGlyLeuGlyTyrTh 866
765 GGACATTCGAGCTACAGATGAGTGTGAGGATATTAGAGCTTTATAG 814
866 rProMetGlyGlyValSerMetTyrValGluThrIleValLysAsnIleL 883
815 ACGAGCTCAACTCGGAAGTGTGCTT..... 840
883 euSerSerAsnSerThrProSerLeuGluArgThrGlyGlnLeuGlyAsp 899
841 .....CCCTCAGAGCTGAAGCACAGTGGAGGAGCTCTACTT 878
900 ValMetLysGluSerSerGluIleSerTyrSerPheSerLysSerPheLe 916
879 A.....CGGACTTCGA 889
916 uSerLysHisPheProAsnAsnLysPhePheGluHisAlaArgLeuHisM 933
890 GTTGGCCACCATAGAGCTTTGAGCATATGAGAGAGAGGACGAGGAAAC 939
933 eHisCysProGluGlySerIle.....SerLysAspGlyProSer 946
940 GCAGACTTAATATG..... 954
947 AlaGlyIleThrMetAlaThrSerLeuLeuSerLeuAlaLeuAspThrPr 963
955 .....CTGTCTCAATATGAGGGGCGAAGCTCGCCCTT..... 987
963 oValProAlaThrThrAlaMetThrGlyGluLeuThrLeuThrGlyLysI 980
987 ..... 987
980 leLeuArgIleGlyLeuArgGluLysThrValAlaAlaLysLeuSer 996

```

seq_name: SwissProt_40:MDH_BACTC

seq_documentation_block:

ID MDH_BACTC STANDARD; PRT; 312 AA.

AC Q9X4K8;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Malate dehydrogenase (EC 1.1.1.37).

GN MDH.

OS Bacillus thermodenitrificans.

OC Bacteria, Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Geobacillus.

OX NCBI_TaxID=33940;

FN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=00462;

RA Williams R.A.D., Welch S.G., Alawadhi S.A.;

RT "Properties and primary structure of a thermostable L-malate

RL dehydrogenase from 'Bacillus thermodenitrificans'";

CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) -> oxalacetate + NADH.

CC -1- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF114423; AAD28555.1;

DR HSPF; Q27743; ICET.

DR InterPro; IPR001557; L_LDH.

DR InterPro; IPR001252; MDH_actsite.

DR InterPro; IPR000205; NAD_binding.

DR InterPro; IPR001236; ldh.

DR Pfam; PF000056; ldh; 1.

DR Pfam; PF02866; ldh_C; 1.

DR PRINTS; PR00086; LLDHDEGNASE.

DR PROSITE; PS00068; MDH; FALSE_NEG.

KW Oxidoreductase; Tricarboxylic acid cycle; NAD.

FT ACT_SITE 153 153 PROTON-RELAY (BY SIMILARITY).

FT BINDING 156 156 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).

FT ACT_SITE 180 180 PROTON-RELAY (BY SIMILARITY).

SQ SEQUENCE 312 AA; 33682 MW; 68C691C462EFF452 CRC64;

alignment_scores:

Quality: 100.50 Length: 216

Ratio: 0.966 Gaps: 9

Percent Similarity: 48.148 Percent Identity: 22.222

alignment_block:

US-09-886-400-3 x MDH_BACTC ..

Align seg 1/1 to: MDH_BACTC from: 1 to: 312

28 CTCGAGTATGCGGAATCCAAAGAGCGAAATCCCAAAG..... 66

130 MetSerTyrThrValPheLysGluSerGlyPheProLysAsnArgValI 146

67GTCAATAGAGAGCGCATACATCCCACTCATCGAGACAC 103

|||||

```
146 eGlyGlnSerGlyValLeuAspThrAlaArg.....PheArgThrP 160
104 TGATTAAGAGAAATTCCTTTGGGCTCAACATAACGGGCTATACCTTA 153
    : : : : : : : : : : : : : : : : : : : : : : : :
160 heValAlaGlnGlu.....LeuAsnIleSer..... 168
154 AAGTTCCTCCGGAAGGATATATAGACCTCGTTAAAGGGGCATCCGAG 203
    : : : : : : : : : : : : : : : : : : : : : : : :
169 .....ValLysAspValThrGlyPheValLeuGlyGlyHisGlyAs 182
204 TGACCTGATAGATATATCGGACGACGACACGACGCAATA.....C 247
    : : : : : : : : : : : : : : : : : : : : : : : :
182 pAspMetValProLeuValArgTyrSerTyrAlaGlyGlyIleProLeuG 199
248 TCCCCCTCCCTCCGCTTAGCAGAGTAGACACACAAGTTCAGAGATAGG 297
    : : : : : : : : : : : : : : : : : : : : : : : :
199 luLysLeuIleProLysAspArgLeuAspAlaIleValGluArgThrArg 215
298 GAAGTTAAGGAAGAGCTCTTCGAGCTTCTCCAAAGGGA.....TTCG 341
    : : : : : : : : : : : : : : : : : : : : : : : :
216 LysGlyGlyGlyGluIleValAsnLeuLeuGlyAsnGlySerAlaTyrTy 232
342 GCTCCAGAGCTCCCTATGACCCGATATCCCTGCCATCTACTAGAGACA 391
    : : : : : : : : : : : : : : : : : : : : : : : :
232 rAlaProAlaAlaSerLeuAlaGluMetValGluAlaIleValLysAspG 249
392 ACGGTTATGATATCTATTCGCGACGGGAGGCGATGCTTTTCTCAGCT 441
    : : : : : : : : : : : : : : : : : : : : : : : :
249 In..... 249
442 CATCTCAACTCGCGATAAAGCCAATTAACCGCTCTATCCACACCTTAT 491
    : : : : : : : : : : : : : : : : : : : : : : : :
250 .....ArgArgIleLeuProAlaIleThrTyrLeu.. 259
492 AAAGCCCCAAGGGAAGGCCCTTAGGTAC.....A 523
    : : : : : : : : : : : : : : : : : : : : : : : :
260 .....GluGlyGluTyrGlyTyrGluGlyIleTyrLeuGlyV 272
524 TCAGCTATCTCCTTGGTCTCAGGAGCTTAGGAAGCGCATAAAGCTCGTT 573
    : : : : : : : : : : : : : : : : : : : : : : : :
272 alProThrIleLeuGlyGlyAsnGlyIleGluLysValIleGluLeuGlu 288
574 TTTGAAGGTAAAGTAAAGCGCTAAAGCGAGTCAAGACATCGAGCCGTA 621
    : : : : : : : : : : : : : : : : : : : : : : : :
289 LeuThrGluAspGluLysAlaAlaLeuAlaLysSerLeuGluSerVal 304
```


OM of: US-09-886-400-3 to: Issued_Patents_AA: * out_format : pfs
Date: Jun 11, 2002 10:34 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters: -DEV=xlh
-MODE=frameat n2p.model -DEV=xlh
-Q/cgn2_1/USPTO_Spool/US09886400/runat_11062002_150914_17298/app_query.fasta_1.1169
-DB-Issued_Patents_AA -QMT=fastan -SUFFIX=ra1 -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.500
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09886400 -CGNJ_1_31 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-886-400-3
Query length: 1095
Database: Issued_Patents_AA: *
Database sequences: 231628
Database length: 2442594
Search time (sec): 25.060000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-613-220B-4 +		1580.00	3187.30	1.4e-170	346
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-07-894-212A-8 +		141.50	250.84	2.7e-07	647
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-07-894-212A-2 +		141.50	250.81	2.7e-07	649
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-07-893-938A-1 +		141.50	250.79	2.7e-07	650
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-270-013B-2 +		104.00	182.69	0.0033	329
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-838-418-2 +		104.00	182.69	0.0033	329
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-09-228-968-72 -		93.00	147.96	0.1030	910
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-09-061-709-2 -		92.50	144.19	0.1331	1142
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-936-135-4 +		91.50	132.27	0.2717	2584
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-418-893D-23 +		88.50	128.73	0.5286	2089
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-418-893D-24 +		88.50	128.73	0.5286	2089
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-09-355-166-1 +		87.50	140.71	0.3617	657
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-09-141-135-2 +		87.00	152.15	0.2332	235
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-248-021A-2 +		86.50	138.10	0.4822	689
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-910-923-3 +		86.00	136.17	0.5729	743
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-09-235-451-36 +		85.50	134.81	0.6629	784
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-907-166-6 +		85.50	134.44	0.6744	788
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-993-118-7 -		84.00	126.89	1.23	1142
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-845-528C-7 -		84.00	126.89	1.23	1142
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-07-718-533-3 +		83.50	132.87	1.01	641
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-161-993-3 +		83.50	132.87	1.01	641
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-487-480A-5 +		83.00	121.22	1.73	1676
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-849-480A-5 +		83.00	134.71	1.01	506
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-09-564-805-237 +		83.00	131.94	1.15	636
/cgn2_6/ptodata/1/iaa/6A_COMB.ppt:US-08-995-280C-2 -		82.50	140.52	0.8454	288
/cgn2_6/ptodata/1/iaa/6A_COMB.ppt:US-09-215-042-2 -		82.50	140.52	0.8454	288
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-676-279-59 +		82.50	132.71	1.21	549
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-804-871-1 +		82.50	132.60	1.21	554
/cgn2_6/ptodata/1/iaa/6A_COMB.ppt:US-09-002-298-1 +		82.00	138.89	0.9508	303
/cgn2_6/ptodata/1/iaa/6B_COMB.ppt:US-09-104-068-4 +		82.00	130.16	1.48	623
/cgn2_6/ptodata/1/iaa/6B_COMB.ppt:US-09-104-068-2 +		82.00	129.89	1.50	637
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-489-733-3 +		81.50	132.02	1.47	491
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-993-581B-3 +		81.50	132.02	1.47	491
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-451-715A-6 +		81.50	126.02	1.94	806
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-09-412-545-2 +		81.00	121.62	2.58	1065
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-484-493-2 +		80.50	128.61	2.04	550
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-484-494-2 +		80.50	128.61	2.04	550
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-345-212-2 +		80.50	128.61	2.04	550
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-09-249-003-2 +		80.50	128.61	2.04	550
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-468-558-2 +		80.00	122.25	2.97	855

/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-676-444-2 + 80.00 122.25 2.97 855
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-118-441-29 + 80.00 119.09 3.43 1110
/cgn2_6/ptodata/1/iaa/6A_COMB.ppt:US-08-338-579A-29 + 80.00 119.09 3.43 111
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.ppt:PCT-US94-09851-29 + 80.00 119.09 3.43 831
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-073-384C-5 + 79.50 121.58 3.33 831

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-613-220B-4

seq_documentation_block:

: Sequence 4, Application US/08613220B
: Patent No. 5958751
: GENERAL INFORMATION:
: APPLICANT: Murphy, Dennis
: APPLICANT: Reid, John
: TITLE OF INVENTION: ALPHA-GALACTOSIDASE
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson, P.C.
: STREET: 4225 Executive Square, Suite 1400
: CITY: La Jolla
: STATE: CA
: COUNTRY: US
: ZIP: 92037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/613,220B
: FILING DATE: 08-MAR-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Haile, Ph.D., Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 09010/004001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-678-5070
: TELEFAX: 619-68-5099
: TELEX:
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 346 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
: US-08-613-220B-4

alignment_scores:
Quality: 1580.00 Length: 364
Ratio: 4.566 Gaps: 18
Percent Similarity: 95.055 Percent Identity: 94.505

alignment_block:

US-09-886-400-3 x US-08-613-220B-4 ..
Align seg 1/1 to: US-08-613-220B-4 from: 1 to: 346
1 TTGAGAGCGCTCGTCTTTCCACGGCAACCTCCAGTATGCGGAATCCCAA 50
|||||
1 LeuArgAlaLeuValPheHisGlyAsnLeuGlnTyAlaGluileProly 17
51 GAGCGAAATCCCAAGGTGATAGAGAAGCATACATCCAGTCATCGAGA 100
|||||
17 sSerGlu...ProLysValleGluLysAlaTyriLeProValleGluT 33
101 CACTGATTAAAGAAGAAATTCCTTTTGGCTCAACATACGGGCTATACC 150
|||||

33 hrLeuIleLysGluGlu...ProPheGlyLeuAsnIleThrGlyTyrThr 48
151 TTAAGTTCCTCCGGAAGGATATATTAGACCTCGTTAAAGGGGGCATCGC 200
49 LeuLysPheLeuProLysAspIleIle...LeuValLysGlyIleAl 64
201 GAGTGACCTGATAGATATCGGACGAGCTACAGCGCAATACCTCC 250
64 aSerAspLeuIleGluIleGlyThrSerTyrThr...AlaIleLeuP 80
251 CCTCCTCCGCTTAGCAGAGTAGAGACACAAGTTCAGAGATAGGAA 300
80 roLeuLeuProLeuSerArgValGluAlaGlnValGlnArgAspArg... 95
301 GTAAGAAGAGAGCTCTCGAGCTTCTCCAAAGGGATTCTGCTGCCAGA 350
96 ValLysGluGluLeuPheGluValSerProLysGlyPheTrpLeuProG 112
351 GCTCGCTATGACCCGATATCCCTCCCTACTGAGGACACGGTTATG 400
112 uLeuAla...AspProIleIleProAlaIleLeuLysAspAsnGlyTyrG 128
401 AGTATCTATTTCGCGGAGGGGAGGAGCTTTCTCAGCTCATCTCAAC 450
128 luTyrLeuPheAlaAsp...GluAlaMetLeuPheSerAlaHisLeuAsn 143
451 TCGGCGATAAGCCAAATTAACCGCTCTATCCACACCTTATAAGGCCA 500
144 SerAlaIleLysProIleLysProLeu...ProHisLeuIleLysAlaG 159
501 AAGGAAAGCGGTTTAGTACATCAGCTATCTCCTTGCTCAGGGAGC 550
159 nArgGluLysArgPheArgTyrIleSerTyrLeuLeu...LeuArgGluL 175
551 TTAGGAAGCGGATAAGCTCGTTTGAAGGTAAGCTAACGCTAAAGGCA 600
175 euArgLysAlaIleLysLeuValPheGluGlyLysValThrLeuLys... 190
601 GTCAAAGACATCGAAGCCGACCGCTTTGGTGGCGGTGAACACGGCTGT 650
191 ValLysAspIleGluAlaValProValTrpValAlaValAsnThrAlaVa 207
651 AATCGCTCGGATCGGAGGCTCTCTTATGATCTCTAAGAAAGTGCGA 700
207 lMetLeu...IleGlyArgLeuProLeuMetAsnProLysLysValAlaAs 223
701 GCTGGATAGGACAGGACCAACATCTTCTATACGGCACCGCATATAGAG 750
223 erTrpIleGluAspLys...AsnIleLeuLeuTyrGlyThrAspIleGlu 238
751 TTCATTGGCTATAGGACATTCGAGCTACAGATGATGATGTTGAGGGATT 800
239 PheIleGlyTyrArgAspIleAlaGly...ArgMetSerValGluGlyLe 254
801 ATTAGAGTTATAGCAGAGCTCAACTCGGAACGTGCCTTCCCTCAGAGC 850
254 uLeuGluValIleAspGluLeuAsnSerGluLeuCys...ProSerGluL 270
851 TGAAGCACAGTGGAGGAGGCTCTACTTACGAGCTTCGAGTTGGGCACCA 900
270 euLysHisSerGlyArgGluLeuTyrLeuArgThrSerTrpAla... 285
901 GATAGAGCTTTGAGGATATGAGAGGAGCGAAGGGAAGCAAGCAAGCTTAA 950
286 AspLysSerLeuArgIleTrpArgGluAspGluGlyAsnAlaArgLeuAs 302
951 TATGCTGTCCTACAAATATGAGGGGCGAAGCTCGGCCCTTTAGCCGAGAACA 1000
302 nMetLeu...TyrAsnMetArgGlyGluLeuAlaPheLeuAlaGluAsnS 318
1001 CGGATGCAAGGGGATGGAGCCCTCCTCCTCAGAGGAGGCTGGATGCTTC 1050
318 erAspAlaArgGlyTrp...ProLeuProGluArgArgLeuAspAlaPhe 333

1051 CGGGCGATATATACGATTCGAGGGGTGAATGGGAACCT 1092
334 ArgAlaIleTyrAsnAspTrpArgGly...AsnGlyGluPro 346
seq_name: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:US-07-894-212A-8
seq_documentation_block:
; Sequence 8, Application US/07894212A
; Patent No. 5368893
; GENERAL INFORMATION:
; APPLICANT: ASADA, KIYOZO
; APPLICANT: UEMORI, TAKASHI
; APPLICANT: MUKAI, HIROYUKI
; APPLICANT: KATO, IKUNOSHIN
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07/894, 212A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95469/C-1195
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-894-212A-8
alignment_scores:
Quality: 141.50 Length: 391
Ratio: 0.773 Gaps: 23
Percent Similarity: 46.803 Percent Identity: 22.762
alignment_block:
US-09-886-400-3 x US-07-894-212A-8 ..
Align seg 1/1 to: US-07-894-212A-8 from: 1 to: 647
73 GAGAAGCATACATCCCGATCGACACTGATTAAGAAGAAATTC 122
27 GluLysCysTyrTrpProPheLeuGluThrLeu.....GluGluTyrPr 41
123 T.....TTTGGGCTCAACATACGGGCTATACCTTAAAGTTCTCC 163
41 oAsnMetLysValAlaIleHisThrSerGlyProLeuIleGluTrpLeuG 58
164 CG.....AAGGATATTATAGACTCGTTAAAGGGGCGCATCGGAGT 204
58 lAspAsnArgProGluTyrIleAspLeuLeuArgSerLeuValLysArg 74

```
205 GACCTGATAGAGATAATCGGAACGAGCTACAGCAGCAATACCTCCCT 254
75 GlyGlnValGluIleValAlaGlyPheTyrGluProValLeuAla 91
255 COTCCGCTTAGCAGAGTAGAAGCAGACAGTTCAGAGAGATAGGAAGTTA 304
91 rtlePro.....LysGluAspArgIleGluGlnIleArgLeuMetL 105
305 AGGAA.....GAGCTCTTCGAGCTTCTCCAAAGGAGTCTGGTCCCA 348
105 yslglutrpAlaLysSerIleGlyPheAspAlaArgGlyValTrpLeuThr 121
349 GAGCTCGCCTATGACCCGATAATCCCTGCCATACCTGAAGACACGGTTA 398
122 GluArgValTrpGlnProGluLeuValLysThrLeuLysGluSerGlyI 138
399 TGAGTATCTATTCGCCGACGGGAGGCGATCTTTCTCAGCTCATCTCA 448
138 eAspTyrValIleValAsp.....AspTyrHisPheM 149
449 ACTCGCGGATAAAGCAATTAACCGCTCTAT...CCACACCTTATAAAG 495
149 etSerAlaGlyLeuSerLysGluLeuTyrTrpProTyrTyrThrGlu 165
496 GCCCAAGGGA.....AAGCGCTTAGSTA 521
166 AspGlyGluValIleAlaValPheProIleAspGluLysLeuArg.. 181
522 CATCAGCTATCTCCTGTGCTCAGGAGCTTAGGAGGCGATAAAGCTC. 570
182 .....TyrLeuIleProPheArgProValAspLysValLeuGluTyrL 196
571 .....GTTTTGAAGTAGGTAGCTAAGCGAGTC.....AAA 606
196 euHisSerLeuIleAspGlyAspGluSerLysValAlaValPheHisAsp 212
607 GACATCGAAGCGGTACCCGCTTGGTGCCCTGAACACGGCTGAATGCT 656
213 AspGlyGluLysPheGlyIleTrpProGlyThrTyrGluTrpValTyr.. 228
657 CGCATCGGAGGCTCTCTATGATCTTAAGAAGTGCAGCTGGA 706
229 .....GluLysGlyTrpL 233
707 TA.....GAGCAACAGGACAATCTT 729
233 euArgGluPhePheAspArgIleSerSerAspGluLysIleAsnLeuMet 249
730 CTATACGGCACCAGATATAG.....TTCATTGGCTATAG 764
250 LeuTyrThrGluTyrLeuGluLysTyrLysProArgGlyLeuValTyrLe 266
765 GGACATTGCGCTACAGAAATGAGTGTGAGGATATTAGAGGTTATAG 814
266 uProIleAlaSerTyr.....PheGluMet.... 274
815 ACAGCTCAACTCGGAGCTGCTCCCTCC..... 843
275 .....SerGluTrpSerLeuProAlaLysGlnAlaArgLeuPhe 287
844 .....TCAGAGCTGAAGCAGCTGGA.....AGGA 869
288 ValGluPheValAsnGluLeuLysValLysGlyIlePheGluLysTyrAr 304
870 GCTCTACTTACGACTTCGAGTTGGCCACAGATAAGAGCTTGAGGATAT 919
304 gValPheValArgGlyGlyIleTrp.....LysAsnPhe...PheT 317
920 GGAGAGGACGACGAGGAC.....GCAAGCTTAATATGCTGTCC 960
317 yrlLysTyrProGluSerAsnTyrMetHisLysArgMetLeuMetValSer 333
```

```
961 TACAATATGAGGGCGAAGTCCGCCCTTTAGCCGAGAAGCAGATCAAG 1010
334 LysLeuValArgAsnAsn..... 339
1011 GGGATGGAGCCCTCCCTCCTGAGAGAGGCTGGATGCCCTCCGGCGATAT 1060
340 .....ProGluAlaArgLysTyrLeuLeuArgAlaGlnC 351
1061 ATACGAT.....TGGAGGGGT 1077
351 yAsnAspAlaTyrTrpHisGly 358
seq_name: /cgr2_5/ptodata/1/1aa/5A_COMB.pep:US-07-894-212A-2
seq_documentation_block:
; Sequence 2, Application US/07894212A
; Patent No. 5366883
; GENERAL INFORMATION:
; APPLICANT: ASADA, KIYOZO
; APPLICANT: UEMORI, TAKASHI
; APPLICANT: MURAI, HIROYUKI
; APPLICANT: KATO, IKUNOSHIN
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABRY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/894,212A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KORULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95469/C-1195
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 649 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-894-212A-2
alignment_scores:
Quality: 141.50 Length: 391
Ratio: 0.773 Gaps: 23
Percent Similarity: 46.803 Percent Identity: 22.762
alignment_block:
US-09-886-400-3 x US-07-894-212A-2
Align seg 1/1 to: US-07-894-212A-2 from: 1 to: 649
73 GAGAGGCTATACATCCAGCTATCGAGACTGATTAAGAGAAATTC 122
29 GluLysCysTyrTrpProPheLeuGluThrLeu.....GluGluTyrPr 43
```

```
123 T.....TTTGGCTCAACATAAGGGCTATACCTTAAAGTTCTCC 163
43 oAsnMetLysValAlaIleHisThrSerGlyProLeuIleGluTrpLeu 60
164 CG.....AAGGATATTATAGACTCGTTAAAGGGGCATCGGAGT 204
60 lAspAsnArgProGluTyrIleAspLeuLeuArgSerLeuValLysArg 76
205 GACGTGATAGATAAATCGAGAGCTACACGACGCATATCTCCCT 254
77 GlyGlnValGluIleValAlaGlyPheTyrGluProValLeuAla 93
255 CTTCCCTTACGAGTAGAGACACAAAGTTTCAGAGATAGGAAAGTTA 304
93 rIlePro.....LysGluAspArgIleGluGlnIleArgLeuMetL 107
305 AGGA.....GACCTCTTCGAGCTTCTCCAAAGGATCTGGCTGCCA 348
107 ySGLuTrpAlaLysSerIleGlyPheAspAlaArgGlyValTrpLeuThr 123
349 GAGCTGCGCTATGACCGGATAATCCCTGCCATACCTGAAGACACGGTTA 398
124 GluArgValTrpGlnProGluLeuValLysThrLeuLysLysSerGlyr 140
399 TGAGTATCTATTCGCCAGCGGGAGGGCGATGCTTTCTCAGCTCATCTCA 448
140 eAspTyrValIleValAsp.....AspTyrHisPheM 151
449 ACTCGGCTATAAGCCCAATTAACCGCTCTAT...CCACACCTTATAAAG 495
151 eSerAlaGlyLeuSerLysGluGluLeuTyrTrpProTyrTyrThrGlu 167
496 CCCCAAGGGAA.....AAGCGCTTTAGGTA 521
168 AspGlyGlyGluValIleAlaValPheProIleAspGlyLysLeuArg.. 183
522 CATCAGCTATCTCTGCTCTCAGGAGCTTAGGAGCGGATAAAGTC. 570
184 .....TyrLeuIleProPheArgProValAspLysValLeuGluTyrL 198
571 .....GTTTGAAGGTAAAGTAAAGTAAAGGAGTC.....AAA 606
198 euHisSerLeuIleAspGlyAspGluSerLysValAlaValPheHisAsp 214
607 GACATCGAAGCGCTACCGCTTGTGGTGGCGGTGAACACGCTGTATGCT 656
215 AspGlyGluLysPheGlyIleTrpProGlyThrTyrGluTrpValTyr.. 230
657 CGGCATCGAAGGCTTCCTCTATGAATCCTAAGAAAGTGGCGAGCTGGA 706
231 .....GluLysGlyTrpL 235
707 TA.....GAGGACAAAGGACAACATCTT 729
235 euArgGluPhePheAspArgIleSerSerAspGluLysIleAsnLeuMet 251
730 CPATACGCCACCGATATAG.....TTTATTGGCTATAG 764
252 LeuTyrThrGluTyrLeuGluLysTyrLysProArgGlyLeuValTyrLe 268
765 GCACATTCGAGCTACAAATGAGTGTGAGGATATTAGAGGTTATAG 814
268 uProIleAlaSerTyr.....PheGluMet... 276
815 ACGAGCTCAACTCGGAAGTGTCCCTTCC..... 843
277 .....SerGluTrpSerLeuProAlaLysGlnAlaArgLeuPhe 289
844 .....TCAGAGCTGAACGACAGTGA.....AGGGA 869
290 ValGluPheValAsnGluLeuLysValLysGlyIlePheGluLysTyrAr 306
870 GCTCTACTACGAGCTTCGAGTGGCCACCAAGATAAGAGCTTGAGGATAT 919
```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-893-928A-1

seq_documentation_block:
; Sequence 1. Application US/07893928A
; Patent No 5578479
; GENERAL INFORMATION:
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: a-AMYLASE FROM HYPERTHERMOPHILIC
; TITLE OF INVENTION: ARCHAEBACTERIUM
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/893,928A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95470/C-1197
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-893-928A-1

alignment_scores:
Quality: 141.50 Length: 391
Ratio: 0.773 Gaps: 23
Percent Similarity: 46.803 Percent Identity: 22.762

alignment_block:
US-09-886-400-3 x US-07-893-928A-1 ..
Align seg 1/1 to: US-07-893-928A-1 from: 1 to: 650

```
73 GAGAAAGCATATCCAGCTATGAGACACTGATTAAGAAGAAATTC 122
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29 GluLysCysTyrTrpPheLeuGluThrLeu.....GluLutyrPr 43
123 T.....TTGGCTCAACATAAGCGGTATACCTTAAAGTTCTCCTC 163
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
43 oMetLysValAlaIleHisThrSerGlyProLeuIleGluTrpLeuG 60
164 CG.....AAGGATATTAGACCTCGTTAAAGGGGCATCGCGAGT 204
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
60 InAsnArgProGluLutyrIleAspLeuLeuArgSerLeuValLysArg 76
205 GACCTGATAGATAATCGAAGAGCTACAGCAGCAATATCTCCCTCCT 254
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
77 GlyInValGluIleValAlaGlyPheTyrGluProValLeuAlaLase 93
255 COTCCGCTTAGCAGAGTAGAAGACACAGTTTCAGAGAGATAGGAGTTA 304
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
93 rIlePro.....LysGluAspArgIleGluGluIleArgLeuMetL 107
305 AGGA.....GAGCTCTCGAGCTTTCTCCAAAGGATTCGCTGCCA 348
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 ysGluTrpAlaLysSerIleGlyPheAspAlaArgGlyValTrpLeuThr 123
349 GAGCTCGCTATGACCGCATATCCCTGCCATAGTGAAGGACACGGTTA 398
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
124 GluArgValTrpGluProGluLeuValLysThrLeuLysGluSerGlyI 140
399 TGAGTATCTATTCCGCGACGGGAGGCGATCTTTCTCAGCTCATCTCA 448
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
140 eAspTyrValIleValAsp.....AspTyrHisPheM 151
449 ACTCGCGATTAAGCCATTAACCGCTCTAT...CCACACCTTATAAG 495
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 eSerAlaGlyLeuSerLysGluGluLeuTyrTrpProTyrThrGlu 167
496 GCCAAAGGGAA.....AAGCGCTTTAGGTA 521
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
168 AspGlyGluValIleAlaValPheProIleAspGluLysLeuArg.. 183
522 CATCAGCTATCTCTGTGTCAGGAGCTTAGNAGCGGATAAGCTC. 570
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 .....TyrLeuIleProPheArgProValAspLysValLeuGluTyrL 198
571 .....GTTTGTGAAGTAAAGCTAAAGGCAGTC.....AAA 606
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 euHisSerLeuIleAspGlyAspGluSerLysValAlaValPheHisasp 214
607 GACATCGAAGCGTACCCGTTGGTGGCGGTGACACAGCGCTGTAAGCT 656
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
215 AspGlyGluLysPheGlyIleTrpProGlyThrTyrGluTrpValTyr.. 230
657 CGCATCGGAAGCTTCCCTTATGAATCCTAAGAAAGTGGCGAGCTGGA 706
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
231 .....GluLysGlyTrpL 235
707 TA.....GAGCAAGACAAACATTCCT 729
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
235 euArgGluPhePheAspArgIleSerSerAspGluLysIleAsnLeuMet 251
730 CTATACGGCACCATATAG.....TTCATTGGCTATAG 764
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 LeuTyrThrGluTyrLeuGluLysTyrLysProArgGlyLeuValTyrLe 268
765 GGACATTGAGGCTACAGATGAGTGTGAGGGATTATTAGAGTTATAG 814
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
268 uProIleAlaSerTyr.....PheGluMet.... 276
815 ACGAGCTCAACTCGGAAGTGTGCTTCC..... 843
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
277 .....SerGluTrpSerLeuProAlaLysGlnAlaArgLeuPhe 289
844 .....TCAGAGCTGAAGCACAGTGA.....AGGA 869
```

```
290 ValGluPheValAsnGluLeuLysValLysGlyIlePheGluLysTyrAr 306
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
870 CTTCTACTTAGGACTTCGAGTTGGCCACCAGATAAGAGCTTGAGGATAT 919
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
306 gValPheValArgGlyIleTrp.....LysAsnPhe...PheT 319
920 GGAGAGAGGACGAAGGGAAC.....GCAGAGACTTAATATGCTGCTCC 960
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
319 yLysTyrProGluSerAsnTyrMetHisLysArgMetLeuMetValSer 335
961 TACATATGAGGGCGAACTCGCCCTTTTACCGGACAGACCGGATGCAAG 1010
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
336 LysLeuValArgAsnAsn..... 341
1011 GGGATGGGAGCCCTCCCTGAGAGAGGCTGGATCGCTTCGGCGCATAT 1060
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 .....ProGluAlaArgLysTyrLeuArgAlaGlnC 353
1061 ATAACGAT.....TGGAGGGGT 1077
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
353 ysAsnAspAlaTyrTrpHisGly 360
seq_name: /cgn2_5/ptodata/1/iaa/5A_COMB.pep.US-08-270-013B-2
seq_documentation_block:
; Sequence 2, Application US/08270013B
; Patent No. 5686294
; GENERAL INFORMATION:
; APPLICANT: Sogabe et al.
; TITLE OF INVENTION: PROTEIN HAVING HEAT-RESISTANT MALATE
; TITLE OF INVENTION: DEHYDROGENASE ACTIVITY
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 61601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/270, 013B
; FILING DATE: 01-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 164701/1993
; FILING DATE: 02-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Robert F.
; REGISTRATION NUMBER: 27555
; REFERENCE/DOCKET NUMBER: 62321
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: (25)3533
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 329 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-270-013B-2
```

```
alignment_scores:
  Quality: 104.00      Length: 154
  Ratio: 1.316        Gaps: 7
Percent Similarity: 51.299 Percent Identity: 25.974
```



```
seq_name: /cgn2_6/ptodata/1/iaa/65_COMB.pep:US-09-228-986-72
seq_documentation_block:
; Sequence 72, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
; FILE REFERENCE: 11000/1020
; CURRENT APPLICATION NUMBER: US/09/228,986
; CURRENT FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 72
; LENGTH: 910
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-228-986-72

alignment_scores:
Quality: 93.00 Length: 388
Ratio: 0.479 Gaps: 25
Percent Similarity: 50.000 Percent Identity: 22.423

alignment_block:
US-09-886-400-3/rev x US-09-228-986-72 ..
Align seg 1/1 to: US-09-228-986-72 from: 1 to: 910
1072 TCCATCGTTATATATCCCGGAGGATCCAGCTCCTCAGGGAGG 1023
200 SerLeuGlnTyrLeuSerLeuSerGln.....GlyAsnLeuGlnGlySe 214
1022 GCTCCCATCCCTTCGATCGCTGTTCTCGCTAAAGGGAGGTCGCC 973
214 rValProSerGluLeuGlyArgLeuSerGlnLeu.....IleVal 227
972 CTCATATGTAGACACATTAAGTCTTCGCTCCTCCTCCCTCTC 923
228 LeuAspLeuPheGlyAsnHisLeuThrGlyCysIle.ProSerSerLeu 244
922 TCCATATC.....CTCAAGCTCTTATCTGGTCCCACTCGAGTCGGT 879
244 erAsnCysThrAsnLeuGlnLeuLeuAsp.IleGlyAspAsnGlnLeuVa 260
878 AAGTAGAGCTCCCTCCACTGTGCTCAGCT.....PheAsnCysThr 848
260 lGlyHisIleProSerHisLeuCysThrLysLysThrThrGlnLeuMetT 277
847 ..CTGAGGAGGACAGTTCGAGTTCGAGTTCGCTCTATATACCTTAATA 800
277 yrLeuArgLeuGly..AlaAsnGlnLeuSerGlySerVal..... 289
799 ATCCCTACACTCATCTCTGAGCTCGCAATGTCCTATACCAATGAAC 750
290 ...ProSerSerLeu.....PheAsnCysThrLysLeuGln 300
749 TCTATATCGGTCCGCTATAGAGAATGTTCCTTCCTCTATCCAGCT 700
301 GluIleAlaLeuProTyrAsnGlnLeuSerGlyIleValProMetGluLe 317
699 CGGC.....ACTTCTTAGGA.....TTCA 680
317 uGlyLysLeuThrHisLeuGlnArgLeuPhePheGlyAsnTyrPheI 334
679 TAAGAGA...AGCTTCGGATCGGACCATACAGCGCTGTTCCAGGCC 633
334 leSerGlyAsnThrMetArgCysProIleLeuThrAlaLeuSerAsnCys 350
632 ACCAAACGGGTACGGCTTCGATGCTTTGACTGCCTTTAGCGTTACCTT 583
```

```
351 SerAspLeuGlnTyrValAspLeuSerGluAsnAsnPheThrGlyGlnLe 367
582 ACCT.....TCAAAAACGAGCTTTATCGCTTCCTCTAA 551
367 uProPheSerIleGlyHisLeuSerLys..LysLeuTyrHisLeu...As 382
550 GTCCTCCTGAGACCAAGGAGATAGCTGATGACCTAAAGCGCTTTTCCCTT 501
382 pleuGlySerAsnGlnLeuAlaGlyGluIleProAlaIleGlyAsnL 399
500 TGGGCTTTATAAGGTGTGATAGAGCGGTTTAATGGCTTTATCGCCGA 451
399 euser..... 400
450 GTTGAGATGAGCTGAGAAAACATCGCTCCCGTCGGCGAATAGACT 401
400 ..... 400
400 CATAACCGTTGCTTCCTCAGTATGCGAGGATATATCGGTTCATAGCGGAGC 351
401 SerLeuThrPheLeuAsnLeuGlyArgAsnTyr...PheThrGlySerI 416
350 TCTGCGACGACGAATCCCTT.....TGGAGAAAGCTCGAAGAGCTCTTC 307
416 le.....ProSerSerLeuIleMetLeuGlnLysLeuGluArgLeuTyr 430
306 CTT.....AACTTCCCTATC..... 292
431 MetAspSerAsnAsnLeuGlnGlyAsnIleProMetGluIleGlyGlnLe 447
291 ....TCTCTGAACCTTGTCTTCTACTCTCTGCTAAGCGGAGGAGGG.... 250
447 uLysSerLeuGlyLeuLeuTyrLeuSerGlyAsnAsnLeuSerGlyLysI 464
249 .....GATATTGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 211
464 leProAspPheValAlaAsnLeuGlnGlnLeuArgTyrLeuTyrLeuAsn 480
211 ..... 211
481 HisAsnGlnLeuSerGlyAspLeuAlaAsnLeuGlyLysCysValas 497
210 .CAGTCTACTCGGATGCGCTTACGAGGTCTATAATATCCTTCGGG 162
497 nLeuLeuLeuAspLeuSerTyrAsnLysLeuSerGlyHisIleProg 514
161 AGGAACCTTTAAGGTATAGCGCTTATGTTAGCCCAAGGAATTCCTTC 112
514 lngLLeuAlaGlyLeuAlaAsnLeuAla.....PheTyr 525
111 TTTAAT 106
526 PheAsn 527
seq_name: /cgn2_6/ptodata/1/iaa/66_COMB.pep:US-09-061-709-2
seq_documentation_block:
; Sequence 2, Application US/090617093
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tsung
; APPLICANT: Gure, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
```

```
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 2
; LENGTH: 1142
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-061-709-2

alignment_scores:
  Quality: 92.50      Length: 236
  Ratio: 0.764       Gaps: 12
  Percent Similarity: 51.271  Percent Identity: 24.576

alignment_block:
US-09-886-400-3/rev x US-09-061-709-2
Align seg 1/1 to: US-09-061-709-2 from: 1 to: 1142

1053 CCGAAGGATCCAGCTCTCTCAGGAGGGCTCCCATCCCTTGCAT 1004
|||||:|||||:|||||:|||||:|||||:|||||:
148 ProGluSerIleGlnSerProPhe...GluGlyPheProGlnSerValLe 163
|||||:|||||:|||||:|||||:|||||:|||||:
1003 CGCTGTCCTCGGCTAAAGGGAGGTGCTGCCCTCATATTGTAGACAGC 954
:|||||:|||||:|||||:|||||:|||||:|||||:
163 uGlnIle.ProValSerAlaIleSerSerThrLeuVal.....Ser 177
|||||:|||||:|||||:|||||:|||||:|||||:
953 ATATTAGCTTTCG.....TTCCC 934
|||||:|||||:|||||:|||||:|||||:|||||:
178 IlePheGlnSerSerProGluSerThrGlnSerProPheGluGlyPhePr 194
|||||:|||||:|||||:|||||:|||||:|||||:
933 TTCGTCCTCTCCATATC.....CTCAGCTCTTATCTGTGCCCCAAC 890
|||||:|||||:|||||:|||||:|||||:|||||:
194 oGlnSerProLeuGlnIleProValSerArgSerPheSerSerThrLeuL 211
|||||:|||||:|||||:|||||:|||||:|||||:
889 TCGAAGTCGTAGTAGAGCTCCCTCCACTGTCTCAGCTCTCAGGGA 840
|||||:|||||:|||||:|||||:|||||:|||||:
211 euserIlePheGlnSerSerProGluArgSerGlnArgThrSerGluGly 227
|||||:|||||:|||||:|||||:|||||:|||||:
839 AGGCAGTTCGAGTGTAGCTCGTCTATACCTCTAATCCCTCAAC 790
|||||:|||||:|||||:|||||:|||||:|||||:
228 PheAlaGlnSerProLeuGlnIleProValSerSerSerSerSerTh 244
|||||:|||||:|||||:|||||:|||||:|||||:
789 ACTCATCTCTAGCCCTGCAATGTCCCTA..... 762
|||||:|||||:|||||:|||||:|||||:|||||:
244 rLeu.....LeuSerLeuPheGlnSerSerProGluArgT 256
|||||:|||||:|||||:|||||:|||||:|||||:
761 .....TAGCCAAATGAACCTATATATCGTGCCGTAT 732
|||||:|||||:|||||:|||||:|||||:|||||:
256 hrGlnSerThrPheGluGlyPheProGlnSerProLeuGlnIleProVal 272
|||||:|||||:|||||:|||||:|||||:|||||:
731 AGAAGATGTGTCTTCCTGCTATATCCAGCTC..... 699
|||||:|||||:|||||:|||||:|||||:|||||:
273 SerArgSerPheSerSerThrLeuLeuSerIlePheGlnSerSerProGl 289
|||||:|||||:|||||:|||||:|||||:|||||:
698 .....GCCACTTCTTAGGATTCATAAGAGGAGCCCTCCGATGC 659
|||||:|||||:|||||:|||||:|||||:|||||:
289 uArgThrGlnSerThrPheGluGlyPheAlaGlnSerProLeuGlnIleP 306
|||||:|||||:|||||:|||||:|||||:|||||:
658 CG.....AGCATACAGCCGTGTTCAGGCCACC 630
|||||:|||||:|||||:|||||:|||||:|||||:
306 roValSerProSerPheSerSerThrLeuValSerIlePheGlnSerSer 322
|||||:|||||:|||||:|||||:|||||:|||||:
629 CAACGGGTACGGCTTCGATG.....TCTTTGACTGCTTTAGCGTTAC 586
|||||:|||||:|||||:|||||:|||||:|||||:
323 ProGluArgThrGlnSerThrPheGluGlyPheProGlnSerProLeuGl 339
|||||:|||||:|||||:|||||:|||||:|||||:
585 CTTACCTCAAAACAGAGCTTTATCCCTTCCTTAAGCTCCCTG..... 543
|||||:|||||:|||||:|||||:|||||:|||||:
339 nileProValSerSerPheSerSerThrLeuLeuSerLeuPheGlns 356
|||||:|||||:|||||:|||||:|||||:|||||:
542 ..AGACAGGAGATACGTGATGTACCTAAAGCGCTTTTCCCTTTGGGCC 495
|||||:|||||:|||||:|||||:|||||:|||||:
;
356 erSerProGluArg.ThrGlnSerThrPheGluGlyPheProGlnSerPr 372
494 TTTA 491
|||||
372 oLeu 373

seq_name: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-936-135-4

seq_documentation_block:
; Sequence 4, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2584 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-936-135-4

alignment_scores:
  Quality: 91.50      Length: 386
  Ratio: 0.517       Gaps: 20
  Percent Similarity: 45.855  Percent Identity: 23.316

alignment_block:
US-09-886-400-3 x US-08-936-135-4
Align seg 1/1 to: US-08-936-135-4 from: 1 to: 2584

45 CCCAAGAGCGAAATCC...AAAGTCATAGAGAGGCATACATCCAG 91
|||||:|||||:|||||:|||||:|||||:|||||:
523 ProHisGluAlaSerProLeuGlyLeuGlyLeuAsnAlaSerProSe 539
|||||:|||||:|||||:|||||:|||||:|||||:
92 TCATCAGCACACTGATTAAAGAAATCTTTGGGCTCAACATACG 141
|||||:|||||:|||||:|||||:|||||:|||||:
539 rGluArgAlaSerAsnProArgProArgGlyLeuTyr.....G 552
|||||:|||||:|||||:|||||:|||||:|||||:
142 GGCTATACCTTAAAGTTCCTCCGAGGATATTATAGACCTCGTTAAGG 191
|||||:|||||:|||||:|||||:|||||:|||||:
552 lLeuTyrMetGluThrProHisGluCysTyr..... 562
|||||:|||||:|||||:|||||:|||||:|||||:
192 GGCATCGCGAGTGACCTGATAGAGATATCGGAACGAGCTACAGCAGC 241
```



```
563 .....SerAlaAr 565
242 CAATACT...CCCTCTCCGCTTAGCAGATAGACAGACAGTTCAG 288
|||||
565 gGlyThyArgAlaSerProAla..ArgGlyLeuGluGlyLeuLeu 581
|||||
289 AGAGATAGGAAGTTAAGGAAGAGCTCTTCGAGCTTTCTCCAAAGGAT 338
|||||
582 GluThrArgProAlaSerProGlyLeuTyProHisGluProArgGly.. 597
|||||
339 CTGCTGCCAGAGCTCCCTATGACCGGATAATCCCTGCCATCTGAAG 388
|||||
598 ....LeuValAlaLeuGlyLeuTyProArgHisIleSerIleLeu... 611
|||||
389 ACAACGGTTATGATCTATTCGCCGACCGGGAG.....GCG 426
|||||
612 .....GluGlyLeuTyAlaArgGlyThrTyArgCysTySer 624
|||||
427 ATGCTTTCTCAGCTCATCTCACTCGGCGATAAGCCATTAACCGCT 476
|||||
625 GlyLeuTyGlyLeuAsnLeuTySerThrHisArgPro...ArgGlyLe 640
|||||
477 CTATCCACACCTTATA.....AAGGCCAAAGGAAAGCGCT 514
|||||
640 uTyAlaArgGlyIleLeuGluAlaArgGlySerGluArgSerGluArgS 557
|||||
515 TTAGGTACATCAGCTATCTCTGCTCTAGG...GAGCTTAGGAAGCG 561
|||||
657 erGluArgGlyLeuTyIleLeuGluLeuGluSerGluArgMetGluThr 673
|||||
562 ATAAAGCTCGTTTGAAGTAAAGTAAAGCTAAAGGCACTCAACACAT 611
|||||
674 ValAlaLeuProHisGluThrTyArgThrHisArgAlaSerProSerGl 690
|||||
612 CGAAGCGTACCCGTTTGGTGGCCGTGAACACGGCTGTATCTCGGCA 661
|||||
690 uArgAlaLeuAlaIleLeuGluAlaLeu.....AlaLeuTySerGlyL 705
|||||
662 TCGAAGGCTTCCTCTTATGAATCCTAAGAAAGTGGCGAGCTGATAGAG 711
|||||
705 euGlyLeuTyProHisGluSerGluArgAlaLeuAla..... 718
|||||
712 GACAAAGGACAACTCTCTATAGCGCACCGATATAGTTCATTGGCTA 761
|||||
719 .....SerAsnThr..... 721
|||||
762 TAGGACATTCGAGCTACAGATGAGTGTGAGGATTATTAGAGTTA 811
|||||
722 .....TyrArgSerGluArgValAlaLeuLeuGluGlyL 733
|||||
812 TAGACGAGCTCACTCGGAAGTGTGCTCTCCCTCAGAG.....CTGAAG 855
|||||
733 euAsnSerGluArgSerGluArgIleLeuGluSerGluArgGlyLeuAla 749
|||||
856 CACAGTGAAGGAGCTCTACTTACGAGCTTCGAGTTGGGCACACAGATA 905
|||||
750 SerProProHisGluLeuTy.....SerCysTySerMetGluThr 763
|||||
905 GAGCTTGAGGATATGAGAGAG..... 927
|||||
763 rGlyLeuAlaLeuAlaLeuGluGlyLeuTyMetGluThrGlyLeuSerG 780
|||||
928 .....GACGAAGGG 936
|||||
780 luArgGlyLeuTyGlyLeuLeuLeuGluHisIleSerSerGluArgAla 796
|||||
937 AACGCAAGACTTAATATCTCTCTCAATATGAGGGCGAAGTCTCGCCT 986
|||||
797 SerProGlyLeuAsnIleLeuGluThrHisArg.....AlaLe 809
|||||
987 TTTATCCGAGAACAGCGATGCAAGGGATGGAGCCCTCTCCCTGAGAGA 1036
|||||
```

```
809 uAlaSerGluArgSer.....GluArg 817
1037 GCGTGGATGCTTCCGGCGCATATAACGATTGGAGGGTGAATGGG 1086
|||||
817 lYLeuAsnThrTyArgGlyLeuTyThrHis....ArgAlaSerAsnThr 832
|||||
1087 GAACCT 1092
833 ArgPro 834

seq_name: /cnp2_6/ptodata/1/iaa/5A_COMB.psp:US-08-418-893D-23

seq_documentation_block:
; Sequence 23, Application US/08418893D
; Patent No. 5559220
; GENERAL INFORMATION:
; APPLICANT: ROESSLER, PAUL G
; APPLICANT: OHLROGGE, JOHN B
; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
; TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTETRA CRYPTICA
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
; STREET: 1617 Cole Blvd.
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,893D
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
; TELEX:
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; US-08-418-893D-23

alignment_scores:
Quality: 88.50 Length: 358
Ratio: 0.497 Gaps: 21
Percent Similarity: 49.721 Percent Identity: 19.553

alignment_block:
US-09-886-400-3 x US-08-418-893D-23 ..
Align seg 1/1 to: US-08-418-893D-23 from: 1 to: 2089

25 AACCTCCAGTATGCCGAATATCCCAAGAGCGAATCCCAAGGTCATAGA 74
|||||
```

```

867 GGAGCTCTACTTACGAGACTTCGAGTTCGGCCACCACGATTAAGAGCTTGAGGA 916
      | |||||:|||||: ||: |||||:|||||:
1302 gLeuthrLysValArgValAspGlu.....fLeGluthrLysValArgV 1317
      : |||||:|||||: |||||: |||||:
917 TATGGAGAGAGGACGAA...GGGAACGCAAGA.....CTTAATATG 954
      : |||||:|||||: |||||: |||||:
1317 althrValGlnAspGluAspGlySerProArgValValProValArgLeu 1333
      : |||||:|||||: |||||: |||||:
955 CTGCTCTACAAATGATGAGGGCGAA 978
      : |||||: |||||:|||||:
1334 ValAlaSerSerMetGlnGlyGlu 1341

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-418-893D-24

seq_documentation_block:
; Sequence 24, Application US/08418893D
; Patent No. 559220
; GENERAL INFORMATION:
; APPLICANT: ROESSLER, PAUL G
; APPLICANT: OHLROGGE, JOHN B
; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
; TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTELLA CRYPTICA
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: NATIONAL RENEWABLE ENERGY LABORATORY
; STREET: 1617 Cole Blvd.
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,893D
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
; TELEX:
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-08-418-893D-24

alignment_scores:
      Quality: 88.50      Length: 358
      Ratio: 0.497      Gaps: 21
Percent Similarity: 49.721      Percent Identity: 19.553

alignment_block:
US-09-886-400-3 x US-08-418-893D-24

```

Align seg 1/1 to: US-08-418-893D-24 from: 1 to: 2089

```

25 AACCTCCAGTATCGCGAAATCCCAAGAGGAAATCCCAAGGTCTATAGA 74
   :::::::::::::::::::::::::::::
1057 SerPheGlnPheAlaAspValProAlaLysAsp.....ArgValThrAr 1071
   :::::::::::::::::::::
75 GAAGGCATATACCCAGTCTCATCCAGACACTGATTAA.....GAAGAA 118
   :::::::::::::::::::::
1071 gGlnGlyPhePheSerValIleAspAspAlaSerLysPheAlaGlnGlnL 1088
   :::::::::::::::::::::
119 TTCTCT.....TTTGGGTCAACATAACGGGCTCTATACCTTA 153
   :::::::::::::::::::::
1088 euProGluIleLeuAsnSerPheGlySerLysIleAlaGlyAspAlaSer 1104
   :::::::::::::::::::::
154 AAGTCTCTCCGAAGGATATATAGACCTCGTTAAAGGGGGCATCCGAG 203
   :::::::::::::::::::::
1105 LysGluGlyProValAsnValLeuGlnVal.....GlyAlaLeuSerGI 1119
   :::::::::::::::::::::
204 TGACCTG...ATAGAGATAATC.....GGAACGAGCTACACGCGCAA 244
   :::::::::::::::::::::
1119 yAspIleSerIleGluAspLeuGluLysAlaThrSerAlaAsnLysAspL 1136
   :::::::::::::::::::::
245 TACTCCCTCTCTCCCTAGCAGTATAGACACAAAGTTTCAGAGAT 294
   :::::::::::::::::::::
1136 ysLeuAsnMetLeuGlyValArgThrValThrAlaLeuIleProArgGly 1152
   :::::::::::::::::::::
295 AGGAAGTTAAGGAAGAGCTCTTCGAGCTTTCTCCAAAGGATTCTGGCT 344
   :::::::::::::::::::::
1153 LysLys.....AspProSerTyrTyrSerPh 1161
   :::::::::::::::::::::
345 GCCAGAGCTGCC.....TATGACCGATAATCCCTGCCATCTGA 385
   :::::::::::::::::::::
1161 eProGlnCysSerGlyPheLysGluAspProLeu..... 1172
   :::::::::::::::::::::
386 AGGACACGGTTATGAGTATCTATTCGCCGACGGGGAGGCGATCTTTC 435
   :::::::::::::::::::::
1172 ..... 1172
436 TCAGTCAATCTCAACTCGGCGATTAAGCCAAATTAACCGCTCATCCACA 485
   :::::::::::::::::::::
1173 .....ArgArgGlyMetArgProThrPheHisI 1182
   :::::::::::::::::::::
486 CCTTATAAGCCCAAGG...GAAAAGCGCTTT..... 516
   :::::::::::::::::::::
1182 sLeuLeuGluLeuGlyArgLeuGluGluAsnPheAlaLeuGluArgIleP 1199
   :::::::::::::::::::::
517 .....AGGTACATCAGCTATCTCTGCTCGAGGAGCTT... 552
   :::::::::::::::::::::
1199 toAlaValcIlyArgAsnValGlnIleTyrValGlySerGluLysThrAla 1215
   :::::::::::::::::::::
553 ...AGGAAGCGGATAAGCTCGTTTTCAGAGGTAAAGTAAAGCTAAGGC 599
   :::::::::::::::::::::
1216 ArgArgAsnAlaAlaGlnValPhe.....LeuArgAl 1227
   :::::::::::::::::::::
600 AGTCAAGACATCGAAGCGCTACCGCTTTGGTGGCGGTGAACACAGCTG 649
   :::::::::::::::::::::
1227 aIleSerHisThrProGlyLeuThrThrPheSerGlyAlaArgAlaL 1244
   :::::::::::::::::::::
650 TAACTCTCGGATCGGAAGGCTCTCTTATGATCTCAAGAAAGTGGCG 699
   :::::::::::::::::::::
1244 euLeuGlnGlyLeuAspGluLeuGluArgAlaGlnAlaAsnSerLysVal 1260
   :::::::::::::::::::::
700 AGCTGGATAGAGGACAGCAACATCTTCTATACGGC..... 738
   :::::::::::::::::::::
1261 Ser...ValGlnSerSerArgIleTyrLeuHisSerLeuProGluL 1276
   :::::::::::::::::::::
739 ACCGATATA.....GAGTTCATTGGCTATAGG 786
   :::::::::::::::::::::
1276 nSerAspAlaThrProGluGluIleAlaLysGluPheGluGly..... 1290
   :::::::::::::::::::::
767 ACATTCCAGGCTACAGATGAGTCTTGAGGGATTATTAGAGTTATAGAC 816
   :::::::::::::::::::::
1291 .....ValIleAsp 1293

```

```

817 GAGCTCAACTCGGAAGTGTGCTTCCTCCTCAGAGCTGAACACAGTGGAG 866
   :::::::::::::::::::::::::::::
1294 LysLeuLysSerArgLeu.....AlaGlnAr 1302
   :::::::::::::::::::::
867 GGAGCTCTACTTACGGACTTCGAGTTGGGCACGACAGATAAGAGCTTGAG 916
   :::::::::::::::::::::
1302 gLeuThrLysLeuArgValAspGlu.....IleGluThrLysValArgV 1317
   :::::::::::::::::::::
917 TATGGACAGAGGAGAGAA...GGGAACGCAAGA.....CTTAATATG 954
   :::::::::::::::::::::
1317 alThrValGlnAspGluAspGlySerProArgValValProValArgLeu 1333
   :::::::::::::::::::::
955 CTGTCCTACATATATAGGGCGCAA 978
   :::::::::::::::::::::
1334 ValAlaSerSerMetGlnGlyGlu 1341
   :::::::::::::::::::::
seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:US-09-355-166-1
seq_documentation_block:
; Sequence 1, Application US/09355166
; Patent No. 6316241
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes
; FILE REFERENCE: GC511-PCT
; CURRENT APPLICATION NUMBER: US/09/355,166
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Bacillus
; US-09-355-166-1

```

```

alignment_scores:
  Quality: 87.50      Length: 333
  Ratio: 0.557       Gaps: 19
  Percent Similarity: 47.147  Percent Identity: 22.823

alignment_block:
US-09-886-400-3 x US-09-355-166-1 ..

Align seg 1/1 to: US-09-355-166-1 from: 1 to: 657

84 CATCCAGTCTATCGAGACTGTATTAAGAGAAATTCCTTTGGGCTCA 133
   :::::::::::::::::::::
61 HisGlyGluLysArgSerThrAspProArgTyrSer.Pro..... 73
   :::::::::::::::::::::
134 ACATACCGGCTATACCTTAAGTTCCTCCCGAG.....GATATT 174
   :::::::::::::::::::::
74 .....AspGlyArgThrLeuAlaPheIleSerAspArgGluGlyAspAla 88
   :::::::::::::::::::::
175 ATAGACCTC.....GTTAAAGGGGTCATCCGAGTACCTGAT 212
   :::::::::::::::::::::
89 AlaGlnLeuTyrIleMetSerThrGluGlyGluAlaArgLysLeuTh 105
   :::::::::::::::::::::
213 AGAGATA...ATCGACAGGAGCTACAG.....CAAG 241
   :::::::::::::::::::::
105 rAspIleProTyrGlyValSerLysProLeuTyrSerProAspGlyGlu 122
   :::::::::::::::::::::
242 CAATATCTCCCTCTCTCCGCTTAGCAGAGTAGACCAAGTTTCAGAGA 291
   :::::::::::::::::::::
122 erIleLeuValThrIleSerLeuGlyGluSerIleAspAspArg 138
   :::::::::::::::::::::
292 GATAGGGAAGTTAAGAA.....GAGCTCTTCGAGCTTTC 326
   :::::::::::::::::::::
139 GluLysThrGluGlnAspSerTyrGluProValGluValGlnGlyLeu 155
   :::::::::::::::::::::
327 TCCAAAG.....GGATTCTGGTGCAGAGCTCGCTATGACCCGATAA 370
   :::::::::::::::::::::

```

```
155 rTyrLysArgAspGlyLysGlyLeuThrArgGlyAlaTyrAlaGlnLeuV 172
371 TCCTGCCATCTGAAG..... 387
172 alLeuValSerValLysSerGlyGluMetLysGluLeuThrSerHisLys 188
388 ...GACAACGGTTATGATGATCTATTCGCGGAGCGGAGCGATCTTT 434
189 AlaAspHisGlyAspProAlaPheSerProAspGlyLysTrpLeuValPh 205
435 CTGAGCTCATCTC.....AACTCGGCGATAAAGCCA..... 465
205 eSerAlaAsnLeuThrGluThrAspAlaSerLysProHisAspValT 222
466 .....ATTAAACGGCTCTATCCACAC... 486
222 yrIleMetSerLeuGluSerGlyAspLeuLysGlnValThrProHisArg 238
487 .....CTTAAAGCCCAAGGAAAGCGCTTTAGGTACATCAGCTA 530
239 GlySerPheGlySerSerPheSerProAspGlyArgTyrLeuAlaLe 255
531 TCCTCTGGTCTCAGGAGCTTAGGAGGCG.....A 562
255 uLeuGlyAsnGluLysGluTyrLysAsnAlaThrLeuSerLysAlaTrpL 272
563 TAAAGCTCGTTTTTGAAGGTAAAGTAACG...CTAAAGCGACGTCAAGAC 609
272 euTyrAspIleGluGlnGlyArgLeuThrCysLeuThrGluMetLeuAsp 288
610 ATCGAACCCGTACCCGTTGGGTGGCGGTGAACACGGCTGTAATGCTCGG 659
289 ValHisLeuAlaAsp.....AlaLeuIleGlyAspSerLeuIleGly 303
660 CATCGGAAGCTTCCTCTATGTAATCCTAAGAAAGTGGCGAGCTGGATAG 709
303 yAlaGluGlnArgProIle.....TrpThrL 312
710 AGNACAAGGACACATCTCTATACGGCACCGATATAGATTCATTGGC 759
312 ysAspSerGlnGlyPheTyrValIleGlyThrAspGlnGlySerThrGly 328
760 TATAGGACATTCAGGCTACAGAATGAGTGTGGAGGATATTAGAGGT 809
329 Ile.....TyrTyrIleSerIleGluGlyLeuValTyrPr 340
810 TATAGACAGCTCAACTCGAAGCTGCGCTTCCTCAGAGCTGAAGCACA 859
340 oIle...ArgLeuGluLysGlu..... 346
860 GTGGAAGGGAGCTCTACTACGAGCTTCGAGTTGGGCACCAAGATAAG 906
347 .....TyrIleAsnSerPheSerLeuSerProAspGlu 357
```

seq_name: /cqn2_6/ptodata/1/1aa/5B_COMB.pep:US-09-141-135-2

seq_documentation_block:

; Sequence 2, Application US/09141135

; Patent No. 5981729

; GENERAL INFORMATION:

; APPLICANT: CHUN, Jong Yoon

; APPLICANT: LEE, Yong Hun

; TITLE OF INVENTION: Transcription Factor Gene Induced by Water Deficit and Abscisic

; FILE OF INVENTION: Acid Isolated from Arabidopsis thaliana

; FILE REFERENCE: 1942/31

; CURRENT APPLICATION NUMBER: US/09/141,135

; CURRENT FILING DATE: 1998-08-27

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: WordPerfect 6.1/Windows

; SEQ ID NO 2

; LENGTH: 235

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-141-135-2

alignment_scores:

Quality: 87.00 Length: 233

Ratio: 0.763 Gaps: 11

Percent Similarity: 48.927 Percent Identity: 21.030

alignment_block:

US-09-886-400-3 x US-09-141-135-2 ..

Align seg 1/1 to: US-09-141-135-2 from: 1 to: 235

493 AAGGCCCAAGGAAAGCGCTTTAGGTACATCAGCTATCTCTGGTCT 542

|||||

27 LysSerAsnAsnGlnLysArgPheAsn..... 35

543 CAGGGAGCTTAGGAGCGCATAAAGCTCTTTTGAAGTAAGTAAGCGC 592

|||||

36 .GluGluGlnIleLysSerLeuGluLeuIlePheGluSerGluThrArgL 52

593 TAAAGGCAGTCAAGACATCGAAGCGGTACCCGTTTGGGTGGCGGTGAAC 642

|||||

52 euLysProArgLysLysValGln..... 59

643 ACGGCTGTAATGTCGGCATCGGAGGCTTCCTCTTATCAATCTTAAGAA 692

|||||

60 ValAlaArgGluLeuGly.....LeuGlnProArgG 70

693 AGTGGGAGCTGGATAGAGGACAAAGGAC..... 720

|||||

70 nMetThrIleTrpPheGlnAsnLysArgAlaArgTrpLysThrLysGlnL 87

721AACATTCTCTATACGGCACCGCATATAGAGTTCATT 756

|||||

87 euGluLysGluTyrAsnThrLeuArg.....Ala 96

757 GGTATAGGACATTCGAGC.....TACAGAATGAGTGTGA 794

|||||

97 AsnTyrAsnAsnLeuAlaSerGlnPheGluIleMetLysLysGluLysG 113

795 GGGATTATTAGAGTTATAGACGAGCTCAACTCGGAACGTGCTTCCTCC 844

|||||

113 nSerLeuValSerGluLeuGlnArgLeuAsnGluMetGlnArgProL 130

845 CAGAGCTGAGCAC.....AGTGAAGGGAGCTCTACTACGG 882

|||||

130 ySGLuLysHisHisGluCysGlyAspGlnGlyLeuAlaLeuSer 146

883 ACTTCGAGTTGGCACCATCAGAT...AAGAGCTTGAGGATATGGAGAGGA 929

|||||

147 SerSerThrGluSerHisAsnGlyLysSerGluProGluGlyArgLeuAs 163

930 CGAAGGGAACGCAAGCTTAATATGCTGCTCCCTACAT..... 966

|||||

163 pGlnGlySerValLeuCysAsnAspGlyAspTyrAsnAsnAsnIleLysT 180

967ATGAGGGC.....GAACTCGCCCTTTTA 990

|||||

180 hrGluTyrPheArgValGlnGlyGluThrAspHisGlnLeuMetAsnIle 196

991 GCGGAGAACACCATCGAAGGGATGGGAGCCCTCCCTGAGAGGAGGCT 1040

|||||

197 ValGluLysAlaAsp..... 201

1041 GGATGCTCTCCGGCGCATATATAACGATTGGGGGTGAAATGGGAA 1089

|||||

202 .AspSerCysLeuThrSerSerGluAsnTrpGlyGlyPheAsnSerAsp 217

seq_name: /cqn2_6/ptodata/1/1aa/5A_COMB.pep:US-08-248-021A-2

seq_documentation_block:

; Sequence 2, Application US/08248021A

```

; Patent No. 5648240
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Jonsson, Klas
; APPLICANT: Patti, Joseph M.
; APPLICANT: Gurudiddappa, Sivashankarappa
; TITLE OF INVENTION: MHC II ANALOG FROM STAPHYLOCOCCUS AUREUS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/248,021A
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: TAMK:155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 689 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-248-021A-2

alignment_scores:
    Quality: 86.50      Length: 271
    Ratio: 0.765        Gaps: 11
    Percent Similarity: 41.697    Percent Identity: 18.819

alignment_block:
US-09-886-400-3 x US-08-248-021A-2

Align seg 1/1 to: US-08-248-021A-2 from: 1 to: 689

97 GAGACACTGATTAAGAAGAAATTCCTTTGGGCTCAACATACGGGCTA 146
||||| :||||| :||||| :||||| :||||| :||||| :|||||
264 GluAlaLysValAsnGlnValProTyrSerIleAsnLeuAsnGlyTh 280
147 TACCTTAAGTTCCTCCCG..... 165
||||| :||||| :||||| :||||| :||||| :||||| :|||||
280 rThrThrAsnIleGlnSerAsnLeuAlaPheSerAsnLysProThrP 297
166 .....AAGCATATTAGACCTCGTTAAAGGGCGCATCGCGAGTAC... 207
||||| :||||| :||||| :||||| :||||| :||||| :|||||
297 snTyrLysAsnLeuThrThrLysValLysSerValLeuLysSerArg 313
207 ..... 207
314 GlyValSerGluArgAspLeuLysHisAlaLysLysAlaTyrThrVa 330
208 .....CTGATAGATATAATCGGAACGAGCT 232
||||| :||||| :||||| :||||| :||||| :||||| :|||||
330 lTyrPheLysAsnGlyGlyLysArgValIleHisLeuAsnSerAsnIle 347
233 ACAGCAGCCCAATACCTCCCTCCCTCCGCTTAGCAGAGTAGAAGACAA 282
||||| :||||| :||||| :||||| :||||| :||||| :|||||
347 yrThrAlaAsnLeuValHisAlaLysAspValLysArgIleGluValThr 363

```

```

283 GTTCAGACAGATAGGGAAGCTTAAGGAAGCTC.....TTCGAGCT 323
||||| :||||| :||||| :||||| :||||| :||||| :|||||
364 ValLysThrValSerLysValLysAlaGluArgTyrValProTyrThr 380
||||| :||||| :||||| :||||| :||||| :||||| :|||||
324 TTCCTCCAAAGGATTCTGGCTGCCAGAGCTC..... 354
||||| :||||| :||||| :||||| :||||| :||||| :|||||
380 eAlaValAsnGlyAlaSerAsnProThrLeuSerAspLeuLysPheThrG 397
355 .....GCCTATGACCCGATA.....ATCCTGTGCCATA 391
||||| :||||| :||||| :||||| :||||| :||||| :|||||
397 lYAspSerArgValSerTyrSerAspIleLysLysLysValLysSerVal 413
382 CTGAAG...GACAAACGGT..... 396
||||| :||||| :||||| :||||| :||||| :||||| :|||||
414 LeuLysHisAspArgGlyIleGlyGluArgGluLeuLysTyrAlaGluL 430
397 .....TATGAGTATCTATTCGCCGACGGGAGGCGATCTTTCTCAG 439
||||| :||||| :||||| :||||| :||||| :||||| :|||||
430 sAlaThrTyrThrValHisPheLysAsnGlyThrLysLysValIle... 445
440 CTCATCTCAACTCGCGATAAGCCAATTAAACGCTCTATCCACACCTT 489
||||| :||||| :||||| :||||| :||||| :||||| :|||||
446 ..AsnLeuAsnSerAsnIleSerGlnLeuAsnLeuLysValLysAsp 461
490 ATAAAG.....GCCCAAAGGGAAGGCGCTT 515
||||| :||||| :||||| :||||| :||||| :||||| :|||||
462 lLeLysAsnIleAspIleAspValLysThrGlyAlaLysAlaLysVal 478
516 TAGGTACATCAGCTAT..... 531
478 rSerTyrValProTyrThrIleAlaValAsnGlyThrThrProIleA 495
532 .....CTCCTTGGTCTCAGGGAGCTT 552
||||| :||||| :||||| :||||| :||||| :||||| :|||||
495 lAsrLysLysLeuLysSerAsnLysGlnLeuIleGlyTyrGlnAspLeu 511
553 AGGAGGCGGATAAAGCTCGTTTGAAGGTAAGTAACGCTAAAGCGAGT 602
||||| :||||| :||||| :||||| :||||| :||||| :|||||
512 AsnLysLysValLysSerValLeuLysHisAsp.....ArgGlyI 525
603 CAAGACATCGAA 615
525 eAsnAspIleGlu 529

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-910-925-3

seq_documentation_block:
; Sequence 3, Application US/08910925
; Patent No. 6162601
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN PININ SPLICE VARIANT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,925
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0365 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 743 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1684847
; US-09-886-400-3 x US-08-910-925-3

```

```

alignment_scores:
  Quality: 86.00      Length: 259
  Ratio: 0.667      Gaps: 11
  Percent Similarity: 49.807  Percent Identity: 22.394

alignment_block:
US-09-886-400-3 x US-08-910-925-3  ..

Align seg 1/1 to: US-08-910-925-3 from: 1 to: 743

229 AGCTACACGACGACATACCCCTCCCTCCGCTTACGAGAGTAGAA... 276
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
514 SerGlnCysHisAlaValLeuGlnSerHisProProSerGlnProGluAs 530
277 .....GCACAAGTTCAGAGATAGGGAAGTTAAGGAAGAGCTCT 316
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
530 pLeuSerLeuAlaValLeuGlnProThrProGlnValThrGlnGlu.... 545
317 TCGAGCTTCTCCAAAGGGATTCCTGCTCCAGAGCTCGCTATGACCG 366
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
546 .....HisGlyHisPheLeuProGluArgLysAspPhePro 557
367 ATA.....ATCCCTGCCACTACGAGCACACGGTTATGAGTA 404
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
558 ValGluSerValLysLeuThrGluValProValAsp.ProValLeuThrV 574
405 TCTAT.....TCGCCAGCGGGAGCGATGCTTTTCTCAGCTC 442
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
574 alHisProGluSerGluSerGluThrAsnThrArg..... 585
443 ATCTCAACTCGCGGATAAAGCCCAATTAAACCGCTCTATCCACACCTTATA 492
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
586 ...SerArgSerArgGlyArgThrArgAsnArgThrThrLysSerArgSe 601
493 AAGCCCAAGGAAAGCTTTAGGTACATCAGCTATCTCTCTGCTCT 542
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 rArgSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 617
543 CAGGGAGCTTAGGAGCGGATAAAGCTCGTTTGAAGTAAGGTAAACCC 592
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
618 SerGlySer.....SerSerSe 623
593 TAAAGGAGTCAAGACATCGAAGCGTACCGCTTTGGGTGGCGGTGAAC 642
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
623 rSerGlySerSerSerSerSerSerSerSerSerSerSerSerSerThrS 640
643 ACGGCTTAATGCTCGGCATCGGAAGGCTTCCTCTTATGAATCCTAAGAA 692
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
640 erGlySerSerArgAspSerSerSerSerSerSerSerSerSerSerGlu 656
693 AGTGGCGAGCTGGATAGAGACAGACATCTCTATACGGCACCG 742

```

```

||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
657 SerArgSerArgSerArgGlyArgGlyHisAsnArgAspArgLysHisAr 673
743 ATATAGAGTTTCATTGCTATAGGACATTGCGAGCTACAGATGAGTCTT 792
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
673 gArgSerValAsp.....ArgLysArgArgAspThrS 684
793 GAGGGATTATTAGAGCTTATAGACGAGCTCAACTCGGAACCTGCTCTCC 842
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
684 erGlyLeuGluArgSerHisLysSerSerLysGlyLysSerSerArgasp 700
843 CTCAGA.....GCTGAAGCACAGTGGAGAGGAGCTCT 874
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
701 ThrLysGlySerLysAspLysAsnSerArgSerArgLysArgSerIl 717
875 ACTTACGGAGCTTCGAGTTGGGCACACAGATAAGAGCTTGAGGATATGGAGA 924
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
717 eSerGluSerSerArgSerGlyLysArg.....SerSerA 729
925 GAGGACGAAGGGAACCGAAGACT 947
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
729 rgSerGluArgAspArgLysSer 736

```

OM of: US-09-886-400-3 to: A_Geneseq_032802.* out_format : pfs
 Date: Jun 11, 2002 10:33 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-WDEL-frame+ n2p.model -DEV=xlh
 -Q/cgn21/USFT0_spool/US09886400/runat_11062002_150913_17285/app_query.fasta.1.1169
 -DB-A_Geneseq_032802 -QFM=fastan -SUFFIX=rag -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62
 -TRANSHUMAN40.cdi -LIST=7 -ALIGN=15 -MODE=LOCAL -OUTEXT=pfs
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTEXT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09886400.@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-886-400-3
 Query length: 1095
 Database: A_Geneseq_032802.*
 Database sequences: 747574
 Database length: 11107396
 Search time (sec): 53.920000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW34643 +	1873.00	3855.49	3.7e-207	36	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	145.50	264.50	2.1e-07	65	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	143.50	260.46	3.6e-07	64	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:AAW34643 +	124.00	211.28	9.4e-05	13	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	121.50	214.74	0.0001	65	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:AAW34643 +	104.50	186.48	0.0086	36	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	104.50	185.97	0.0088	37	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	104.50	186.50	0.0094	32	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:AAW34643 +	104.00	180.01	0.0124	57	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	104.00	180.01	0.0124	57	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	104.00	177.55	0.0138	70	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	104.00	177.55	0.0138	70	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	104.00	177.19	0.0140	72	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	104.00	177.19	0.0140	72	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	103.50	173.16	0.0182	93	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	103.50	172.18	0.0190	10	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	103.50	166.74	0.0241	16	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	97.00	159.76	0.1023	92	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	96.50	159.78	0.1117	84	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:AAW34643 +	96.00	155.99	0.1437	10	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	96.00	155.21	0.1487	11	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	94.50	163.94	0.1327	41	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	94.00	163.78	0.1459	38	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	93.50	165.15	0.1502	31	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	93.00	151.71	0.2335	91	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	93.00	144.58	0.2397	16	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	93.00	140.31	0.2821	23	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	92.50	148.20	0.3764	41	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	92.50	148.00	0.3766	11	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	92.00	166.55	0.1841	21	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	92.00	166.45	0.1850	21	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	92.00	165.56	0.1922	23	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	92.00	155.25	0.3002	56	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:AAW34643 +	92.00	144.42	0.4799	14	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	91.50	159.38	0.2744	36	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	91.50	156.29	0.3135	47	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +	91.50	156.29	0.3135	47	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	91.00	160.58	0.2845	30	
/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +	91.00	159.82	0.2940	32	

/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW80176 + 91.00 159.82 0.2940
 /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW82559 + 91.00 159.82 0.2940
 /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW82373 + 91.00 157.99 0.3182
 /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW62551 - 91.00 152.72 0.3998
 /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW51656 + 91.00 148.33 0.4833

seq_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW34643

seq_documentation_block:

ID AAW34643 standard; Protein: 364 AA.
 AC AAW34643;
 DT 27-MAR-1998 (first entry)
 DE Thermostable alpha-galactosidase AEDII12RA-alpha-gal-18GC.
 KW Alpha-galactosidase; alpha-glycosidase; thermostable enzyme;
 KW food processing; alpha glycoside hydrolysis; raffinose;
 KW stachyose; verbascose; bean; flatulence; AEDII12RA-alpha-gal-18GC.
 OS Thermococcus alcaliphilus strain AEDII12RA.

XX Key Location/Qualifiers

FT Misc-difference 329 /note= "encoded by CTT"

XX WO9732974-A1.

XX 12-SEP-1997.

XX 05-FEB-1997; 97WO-US01452.

XX 08-MAR-1996; 96US-0613220.

XX (RECO-) RECOMBINANT BIOCATALYSIS INC.

XX Murphy D, Reid J, Rudolph MJ;

XX WPI: 1997-470541/43.

XX N-FSDS: AAT93753.

XX Nucleic acid encoding alpha-galactosidase from Thermococcus

XX alcaliphilus - used in food processing to hydrolyse

XX alpha-glycosides, e.g. raffinose

XX Claim 1; Fig 1; 32pp; English.

XX This protein comprises AEDII12RA-alpha-gal-18GC, a claimed
 CC thermostable alpha-galactosidase of Thermococcus alcaliphilus
 CC AEDII12RA, a bacterium that shows optimum growth at 85 deg C and
 CC pH 9.5. Also claimed are: (1) an isolated polynucleotide (see
 CC AAT93753) encoding the alpha-galactosidase; (2) a vector containing
 CC the polynucleotide or homologous or complementary sequences; (2)
 CC host cells containing the vector; (3) a process for producing the
 CC alpha-galactosidase in transformed or transfected host cells; an
 CC enzyme showing at least 70% identity to alpha-galactosidase and
 CC comprising at least 30 amino acid residues of its sequence; and (4)
 CC a method for hydrolysing alpha-galactoside bonds using the enzyme.
 CC The enzyme can be used to hydrolyse raffinose to sucrose and glucose
 CC in sugar beet processing (raffinose inhibits crystallisation of
 CC sucrose), and as a digestive aid to hydrolyse raffinose, stachyose
 CC and verbascose in beans and other gassy foods.

XX Sequence 364 AA;

XX alignment_scores: Quality: 1873.00 Length: 364

XX Ratio: 5.146 Gaps: 0

XX Percent Similarity: 100.000 Percent Identity: 99.725

XX alignment_block:

US-09-886-400-3 x AAB34643 ..

Align seg 1/1 to: AAB34643 from: 1 to: 364

```
1 TTGAGAGCGGTCGCTTTTCACGGCAACCTCCAGTATGCGCAATCCCAA 50
1 LeuArgAlaLeuValPheHisGlyAsnLeuGlnTyAlaGluLeuProly 17
51 GAGGGAATCCAAAGGTCATAGAGAGGAGCATACCCAGTCATCGAGA 100
17 sSerGluileProlyValileGluLysAlaTyriileProvalileglut 34
101 CACTGATTAAAGAAATTCCTTTGGGCTCAACATAACGGGCTATACC 150
34 hrLeuileLysGluGluileProPheleLysasnlethrGlyTyThr 50
151 TTAAGTTCCTCCGAGGATATATAGACTCGTTAAAGGGGCGCATCGC 200
51 LeulysPheLeuProLysAspIleleAspLeuValLysGlyIleAl 67
201 GAGTGACCTGATAGAGATAATCGGAAGAGCTACACGACGCAATACTCC 250
67 aserAspLeuileGluileleGlyThrSerTyThrHisAlaileLeuP 84
251 CCCTCCTCCCGCTTAGCAGAGTAGAAGCACAACTTCAGAGAGATAGGAA 300
84 roLeuLeuProLeuSerArgValGluAlaGlnValGlnArgAspArgGlu 100
301 GTTAGAGAGAGCTCTTCGAGCTTCTCCAAAGGATTCGGCTGCCAGA 350
101 VallyGluGluLeuPheGluLeuSerProLysGlyPheTrpLeuProgl 117
351 GCTCGCTATGACCCGATAATCCCTGCTCATCTGAAGGACAAAGGTTATG 400
117 uLeAlaTyAspProileleProAlaileLeuLysAspAsnGlyTyrg 134
401 AGTATCTATTCGCGAGCGGAGCGGAGTGTCTTCAGCTCATCTCAAC 450
134 LuTyLeuPheAlaAspGlyGluAlaMetLeuPheSerAlaHisLeuasn 150
451 TCGGCGATAAGCAATTAACCGCTCTATCCACACCTTATAAGGCCCA 500
151 SerAlaileLysProileLysProLeuTyProHisLeuileLysAlaGl 167
501 AAGGAAAGCGTTTAGTACATACATCTCCTTGGTCTCAGGAGC 550
167 nArgGluLysArgPheArgTyriileSerTyLeuLeuGlyLeuArgGluL 184
551 TTAGGAAGGCGATAAAGCTCGTTTTGAAGGTAAGGTAAGCTAAAGGCA 600
184 euArgLysAlaileLysLeuValPheGluGlyLysValThrLeuLysAla 200
601 GTCAAAGACATCGAAGCGCTACCCGTTTGGGTGGCGCTGAACACGCTGT 650
201 VallyAspIleGluAlaValProValTrpValAlaValAsnThrAlaVa 217
651 AATCGTCGGATCGAAGGCTTCTCTTATGAATCTTAAGAAAGTGGCA 700
217 lMeLeuGlyIleGlyArgLeuProLeuMetAsnProLysLysValAlas 234
701 GCTGGATAGAGGACAAGCAACATCTTCTATACGGCACCGCATATAGAG 750
234 erTrpIleGluAspLysAspAsnileLeuLeuTyrglyThrAspIleGlu 250
751 TTCATTCGCTATAGGCAATGAGGCTACAGATGAGTGTAGGGGAT 800
251 PheileGlyTyArgAspIleAlaGlyTyArgMetSerValGluGlyLe 267
801 ATTAGAGTTATAGACAGCTCAACTCGGAAGTGTGCTTCCCTCAGAGC 850
267 uLeGluValIleAspGluLeuAsnSerGluLeuCysLeuProSerGluL 284
851 TGAAGCACATGGAGGAGGAGCTCTACTTACGGACTTCGGAGTGGGCACCA 900
```

```
|||||
284 euLysHisSerGlyArgGluLeuTyLeuArgThrSerSerTrpAlaPro 300
901 GATAAGAGCTTGAGGATATGAGAGAGGACGAAGGGAACGCAAGACTTAA 950
|||||
301 AspLysSerLeuArgIleTrpArgGluAspGluGlyAsnAlaArgLeuAs 317
951 TATCTCTCTCTACAATATGAGGGCGGAACCTCGCCCTTTAGCCGAGAACA 1000
|||||
317 nMeLeuSerTyAsnMetArgGlyGluLeuAlaPheLeuAlaGluAsn 334
1001 GCGATGCAAGGGATGGAGCCCTCCCTCGAGAGAGGCTGGATGCTTC 1050
|||||
334 erAspAlaArgGlyTrpGluProLeuProGluArgArgLeuAspAlaPhe 350
1051 CGGCGGATATATACGATTGAGGGGTGAAATGGGGAACCT 1092
|||||
351 ArgAlaileTyAsnAspTrpArgGlyGluAsnGlyGluPro 364
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAB96088
seq_documentation_block:
ID AAB96088 standard; Protein; 655 AA.
XX
AC AAB96088;
XX
DT 29-OCT-2001 (first entry)
XX
DE Putative P. abyssi alpha-amyase.
XX
KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX
PN FR2792651-Al.
XX
PD 27-OCT-2000.
XX
PF 21-APR-1999; 99FR-0005034.
XX
PR 21-APR-1999; 99FR-0005034.
XX
PA (CNRS ) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querallou J, Weissenbach J, Saurin W, Heilig R;
XX
WPI; 2001-126236/14.
XX
New nucleotide sequences isolated from Pyrococcus abyssi encode
proteins useful in industry -
Claim 7; Pages 711-712; 1657pp; French.
XX
The present invention relates to the genomic sequence of Pyrococcus
abyssi (see AAF86431 and AAB41223-7) and P. abyssi proteins. P. abyssi is
a hyperthermophilic archaeon, which is isolated from deep-sea P. abyssi
hydrothermal vents. The present sequence is one such P. abyssi protein.
The proteins of the present invention have various potential industrial
uses, since the proteins are stable at very high temperatures, some up to
110 degrees centigrade.
CC Note: This patent is in the same patent family as WO2000065062, which
contains additional sequences as shown in AAB99132-AAB99143,
AAH75903-AAH75920 and AAG66436.
XX
SQ Sequence 655 AA;
```

```
alignment_scores:
Quality: 145.50 Length: 401
Ratio: 0.808 Gaps: 22
Percent Similarity: 44.888 Percent Identity: 22.444
```

alignment_block:

US-09-886-400-3 x AAB96088

Align seg 1/1 to: AAB96088 from: 1 to: 655

```

73 GAGAGGATACATCCAGTCCAGTCCAGACACTGATTAAAGAAAGAAATCC 122
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
28 GluLysAlaTyrArgProPheLeuGluLeu.....GluGluTyrPr 42
123 T.....TTGGCTCAACATACGGGTATACCTTAAAGTTCCTCC 163
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
42 OasnMetLysValAlaIleHisSerGlyLeuValGluTrpLeuG 59
164 CGAAG.....GATATTAGACCTCGTTAAAGGGGCATCGCGAGT 204
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
59 LuGluAsnLysProAspTyrIleAspLeuLysSerLeuValArgLys 75
205 GACCTGATAGATAATCGGACGAGCTACACGACGCACTACTCCCT 254
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
76 GlyGlnValGluIleValAlaGlyPheTyrGluProValLeuAla 92
255 CTCCTCCGCTTAGCAGATAGAGCAGAGTTTCAGAGAGATAGGAAGTTA 304
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
92 aileProLysGluAspArgLeuGluGlnIleTyrLeuLysGluTrpA 109
305 AGGAAGACTCTTCGAGCTTCTCCAAAGGATTCTGGCTGCCAGAGCTC 354
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
109 laLysLysIle...GlyTyrAspAlaLysGlyLeuTrpLeuThrGluArg 124
355 GCCTATGACCCGATATCCCTGCCATACCTGAAGGACACGCTTATGAGTA 404
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
125 ValTrpGlnProGluLeuValLysThrLeuArgGluAlaGlyIleGluTyr 141
405 TCTATCCCGGAGGGGAGGCGATCTTTCTCAGCTCATCTCACTCGG 454
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
141 rValValValAsp.....AspTyrHisPheMetSerA 152
455 CG...ATAAGGCAATTAACCGCTCATCCACACACTTATAAGGCCCAA 501
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
152 laGlyLeuSerLysAspGlnLeuPheTrpProTyrTyrThrGluAspGly 168
502 AGGGA.....AAGGCCTTAGGTAC..... 522
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
169 GlyGluValIleThrValPheProIleAspGluLysLeuArgTyrLeuI 185
523 .....ATCAGCTATCTCTCTGCTGTCTCAGG 547
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
185 eProPheArgProValAspLysValIleSerTyrLeuHisSerLeuAla 202
548 AGCTTAGGAAGCGGATAAGTCGCTTTTGTAGGTAAGTAACGCTAAG 597
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
202 erGluAspGluSerLysValAlaValPheHis..... 212
598 GCAGTCAAGACATCGAGCCGTCCTGTTGGTGGCGGTGAACACGCG 647
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
213 .....AspAspGlyGluLysPheGlyIleTrp..... 221
648 TGTATGCTCGGCATCGGAAGCTTCTCTTATGAATCCTAAGAAAGTGG 697
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
222 .....ProMet.....Thr 225
698 CGAGCTGATAGGACAGGAC..... 720
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
225 yrGluTrpValTyrGluLysGlyTyrLeuArgGluPheAspArgVal 241
721 .....AACATTCTTATACGGCACCGCATATAGAG... 750
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
242 SerSerAspGluAlaIleAsnIleMetLeuTyrSerGluTyrLeuGlnLys 258
751 .....TTCATGGCTATAGGACATTCAGGCTAC..... 780
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
258 sPheLysProLysGlyLeuValTyrLeuProIleAlaSerTyrPheGlu 275

```

```

781 .....AGATGAGTGTGAGGA 798
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
275 etSerGluTrpSerLeuProAlaGlnGlnAlaLysLeuPheValGlu... 290
799 TTATTAGAGTTATAGACGAGCTCAACTCGGAAGTGTGCTTCCCTCAGA 848
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
291 PheValGluLysLeuLysGluLeuAsn..... 299
849 GCTGAAGCACAGTGAAGGAGCTCTACTTACGGACTTCGAGTTGGGCAC 898
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
300 .....MetPheGluArgTyrArgValPheValArgGlyIleTrp.... 313
899 CAGATAAGAGCTTGAAGTATCGAGAGGACGAGGAGGAC..... 939
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
314 .....LysAsnPhe...PheTyrLysTyrProGluAlaAsnTyrMetHis 327
940 GCAAGACTTAATATGCTCTCTACATAATAGGGGCGAAGTTCGCCCTTTT 989
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
328 LysArgMetLeuMetLeuSer.....ArgLeuLe 337
990 AGCCGAGAACACGCGATGCAAGGGATGGAGCCCTCCTCGAGAGGAGGC 1039
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
337 uArgAspAsnProSerAlaArgPhe..... 346
1040 TGGATGCTTCGGCGCATATATAACGAT.....TGGAGGGGTGAAAAAT 1083
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
347 .....ValLeuArgAlaGlnCysAsnAspAlaTyrTrpHisGlyValPhe 361
1084 GGG 1086
|||
362 Gly 362

```

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:AA47504

seq_documentation_block:

ID AA47504 standard; protein; 649 AA.

```

XX AA47504;
XX AC
XX DT 07-JUL-1994 (first entry)
XX DE Pyrococcus furiosus alpha amylase.
XX KW Pyrococcus furiosus; alpha amylase; liquefaction; polymers;
XX OS Pyrococcus furiosus.
XX PN EP577257-A.
XX PD 05-JAN-1994.
XX PF 17-MAY-1993; 93EP-0303801.
XX PR 09-JUN-1992; 92US-0893928.
XX PA (UIJO ) UNIV JOHNS HOPKINS.
XX PI Anfinson CB, Laderman K;
XX DR WPI; 1994-009532/02.
XX PT Purified Pyrococcus furiosus alpha-amylase - used for the
XX PT industrial liquefaction of gluco-polymers at high temps.
XX PS Claim 2; Figure 9; 4lpp; English.
XX CC The purified pyrococcus furiosus alpha amylase can act on substrates
XX CC with a low degree of polymerisation. e.g. glucose polymers as short
XX CC as maltotriose. The enzyme can be used for efficient industrial
XX CC liquefaction of glucopolymers at high temperatures.

```

SQ Sequence 649 AA;
 alignment_scores:
 Quality: 143.50 Length: 399
 Ratio: 0.776 Gaps: 24
 Percent Similarity: 46.366 Percent Identity: 22.807
 alignment_block:
 US-09-886-400-3 x AAR47504 ..
 Align seg 1/1 to: AAR47504 from: 1 to: 649

```

73 GAGAGGCATACATCCAGTCATCGACACACTGATTAAAGAAATTC 122
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
29 GlulysCysTyrTrpProPheLeuGluThrLeu.....GluGluTyrPr 43
123 T.....TTTGGGCTCAACATAAGGGCTATACCTTAAAGTTCTCTCC 163
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
43 cAsnMetLysValAlaIleHisThrSerGlyProLeuIleGluTriLeu. 59
154 CGAAGGATATT.....ATAGACCTCGTTAAAGGGGCATCGG 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||
60 ..GlnAspIleArgProGluTyrIleAspLeuLeuArgSerLeuValLys 75
202 AGTGACCTGATAGATAATCGGAACGAGCTACACGACGCAATACTCC 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||
76 ArgGlyGlnValGluIleValAlaIleGlyPheTyrGluProValLeuAl 92
252 CCTCTCCCGCTTAGCAGAGTAGAAGCACAAAGTTCAGAGAGATAGGAAG 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||
92 aSerIlePro.....LysGluAspArgIleGluGlnIleArgLeuM 106
302 TTAAGGAA.....GAGCTCTTCGAGCTTTCTCAAAGGATTCTGGTG 345
|||||:|||||:|||||:|||||:|||||:|||||:|||||
106 eLysGluTrpAlaLysSerIleGlyPheAspAlaArgGlyValTrpLeu 122
346 CCAAGCTCGCCTATGACCGGATANTCCCTGCCATACCTGAAGACAACGG 395
|||||:|||||:|||||:|||||:|||||:|||||:|||||
123 ThrGluArgValTrpGlnProGluLeuValLysThrLeuLysGluSerGl 139
396 TTATGAGTATCTATCGCGAGGGGAGGCGATCTTTCTCAGCTCATC 445
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
139 yIleAspTyrValIleValAsp...AspTyrHisPheMetSerAlaGluL 155
446 TCAACTCG.....GCTTTTGAAGCTAAGTA 588
|||||:|||||:|||||:|||||:|||||:|||||:|||||
155 euSerLysGluLeuLeuTyrTrpProTyrTyrThrGluAspGlyGlu 171
454 .....GGATAAAGCCCAATTAAACCGCTCTATCCACACCTTATAAGGC 497
|||||:|||||:|||||:|||||:|||||:|||||:|||||
172 ValIleAlaValPheProIleAsp..... 179
498 CCAAGGGAAGAGCGCTTAGGTACATCAGCTATCTCTTGGTCTCAGGG 547
|||||:|||||:|||||:|||||:|||||:|||||:|||||
180 .....GluLys.....LeuArgTyrLeuIleProPheArgp 190
548 AGCTTAGAGGCGGATAAGCTC.....GTTTGTGAAGCTAAGTA 588
|||||:|||||:|||||:|||||:|||||:|||||:|||||
190 roValAspLysValLeuGluTyrLeuHisSerLeuIleAspGlyAspGlu 206
589 ACAGCTAAAGGAGTC.....AAAGCATCGAAGCCGCTACCCGTTTGGGT 632
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
207 SerLysValAlaValPheHisAspAspGlyGluLysPheGlyIleTrpPr 223
633 GCGCGTGAACAGCGCTGTAATGCTCGGCATCGGAGGCTTCTCTTATGA 682
|||||:|||||:|||||:|||||:|||||:|||||:|||||
223 oGlyThrTyrGluTrpValTyr..... 230
683 ATCCTAAGAAAGTGGCGAGCTGGATA..... 708
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
231 .....GluLysGlyTrpLeuArgGluPheAspArgIleSer 243

```

```

709 ...GAGACAAGGACAAACATCTTCTATATAGGACCGATATAGAG..... 750
:|||||:|||||:|||||:|||||:|||||:|||||:|||||
244 SerAspGluLysIleAsnLeuMetLeuTyrThrGluTyrLeuGluLysTy 260
751 .....TTCAATTGGCTATAGGACATTCAGGCTACAGAATGAGTG 790
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
260 rLysProArgGlyLeuValTyrLeuProIleAlaSerTyr..... 273
791 TTGAGGATTATTAGAGTTATTAGACGAGCTCAACTCGGAAGTGCCTT 840
:|||||:|||||:|||||:|||||:|||||:|||||:|||||
274 .....PheGluMet.....SerGluTrpSerLeu 281
841 CCC.....TCAGAGCTGAAGCA 857
282 ProAlaLysGlnAlaArgLeuPheValGluPheValAsnGluLeuLysVa 298
858 CAGTGA.....AGGAGCTCTACTTACGAGCTTCGAGTTGGG 895
:|||||:|||||:|||||:|||||:|||||:|||||:|||||
298 lLysGlyIlePheGluLysTyrArgValPheValArgGlyGlyIleTrp. 314
896 CACCAGATAAGAGCTTCAGGATATGAGAGAGACGACGAGGGAAC..... 939
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
315 .....LysAsnPhe...PheTyrLysTyrProGluSerAsnTyrMet 327
940 ...GCAAGACTTAAATATGCTGTCTCTACAATATAGGGGCGCAACTCGCCT 986
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
328 HisLysArgMetLeuMetValSerLysLeuValArgAsnAsn..... 341
987 TTTAGCCGACAACAGCATCGAAGGGGATGGGAGCCCTCCCTGAGAGGA 1036
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
342 .....ProGluAla 345
1037 GCGTGGATGCTTCGCGCGATATATAACCAT.....TGGAGGGGT 1077
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
345 rgLysTyrLeuLeuArgAlaGlnCysAsnAspAlaTyrTrpHisGly 360

```

seq.name: /SDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.AA996091

seq_documentation_block:

ID AAB96091 standard; Protein; 1362 AA.

XX AAB96091;

XX AC AAB96091;

XX DT 29-OCT-2001 (first entry)

XX DE Putative P. abyssi amylopullulanase.

XX KW Hyperthermophilic archaeon; hyperthermophilic protein.

XX OS Pyrococcus abyssi.

XX PN FR2792651-Al.

XX PD 27-OCT-2000.

XX PF 21-APR-1999; 99PR-0005034.

XX PR 21-APR-1999; 99PR-0005034.

XX PA (CNRS) CNRS CENT NAT RECH SCI.

XX PI (IFRE-) IFREMER INST FR RECH EXPL MER.

XX PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

XX PI Querellou J, Weissenbach J, Saurin W, Heilig R;

XX DR WPI; 2001-126236/14.

XX PT New nucleotide sequences isolated from Pyrococcus abyssi encode

XX PT proteins useful in industry -

XX PS Claim 7; Pages 715-719; 1657pp; French.

XX CC The present invention relates to the genomic sequence of Pyrococcus

CC abyssi (see AAF8431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. abyssi protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as WO2000065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAH75903-AAH75920 and AAG66436.

AA	Sequence	1362 AA;
SQ		

alignment_scores:		
Quality:	124.00	Length: 381
Ratio:	0.756	Gaps: 16
Percent Similarity:	43.045	Percent Identity: 21.522

```
alignment_block:
```

Align seg 1/1 to: AAB96091 from: 1 to: 1362

[illegible]

427 TrpGluHisTy rPropheAspGlyLysLeuPheLeuGluGluLeuTy rAr 443
606 AGACATCGAAGCCGTACACCCCTTTGGTGGCCGTGAACAGCGCTGA.... 651
::: :::
443 gGlnLeuGluLeuGlnLysTy sGlyLeuIleArgThrValTr pRos 460
652ATGCCTCGGC.....ATCGGAAGCCTTCCTCTTATG 681
||| :::
460 erGluTy rIleGluMetPheGlyAspLysAlaAsnLysLeuThrPr oLys 476
682 AATCCTTAAGAAAGTGCGCAGCTGGATGAGGACAAGCAACATTCTTCT 731
||| :::
477 MetMetLysArgLeuAspPheThrThrGluAspAsnValAsnAlaLeuLe 493
732 ATACGCCACCGATATAGAGTTCATNTGCTATAGGGACATTCAGGCTACA 781
|| :::
493 uLYsAlaLysThrLeu.....GlyGluL 501
782 GAATGACTGTGTAGGGATTATTAGAGGTTATACAGAGCTCAACTCGGAA 831
::::: ||| :::
501 euTy rAspMetValGlyValThrGlu..... 509
832 CTGTGCCTTCCTCAGAGCTGAAGCAGTGGGAAGGGAGCTCTACTTACG 881
||| :::
510GluMetGlnTrpPr 514
882 GACTTCGAGTGTGGCACACAGATAAGAGCTTGAGGATATGGAGAGAGGACG 931
||||| :::
514 ogLusErserTriple...AspGlyThrLeuSerThrTrp.....I 527
932 AAGGGAAGCGAAGACTTAATATGCTGCTCTACAATATGAGGGCGAACCTC 981
||||: :::
527 leGlyGluProGInGluAsnIleAlaTrpTy rTrpLeuTy rLeuAlaArg 543
982 GCCCTTTTAGCCGAGAACAGCAT...GCAAGGGGATGGAGCCCCCTCCC 1028
||||| :::
544 LysAlaLeuPheGluAsnLysAspAsnValLysAspTrpAsnLysAlaty 560
1029 TGAGAGGAGCGCTGGATGCCCTCCGGGGGATATATACGATTGG 1071
||| |||
560 rGLuTy r.....LeuPheArgAlaGluGlySerAspTrp 571

seq name: /SIDSl/acadata/geneseq/geneseq-emb1/AA1998.DAT:AAW54870

seq_documentation_block:

ID AAW54870 standard; Protein; 653 AA.

AC AAW54870;

DT 01-SEP-1998 (first entry)

DE Super heat resistant 4-alpha-glucanotransferase.

Super heat-resistant 4- α -glucanotransferase; heat-treatment;
KW
KW
KW alpha-1,4-glucan; alpha-1,4-glucoside bond.

OS Pyrococcus sp.

PN JP10150986-A.

09-JUN-1998.

21-NOV-1996; 96JP-03111117.

AA 21-NOV-1996; 96JP-03111117.
PR

PA (BEAB-) BE ABLE KK.

PA (NAGA-) NAGASE SETRA
XX

DR WPI; 1998-379989/33.
DR N-PSDB: AAV27026.

New 4-alpha-glucosyl

PT high temperatures to transfer at least one glucose unit

XX Claim 4; Page 10-13; 18pp; Japanese.

CC The super heat-resistant 4-alpha-glucanotransferase has an optimum pH of
 CC 6.0-8.0, with an optimum temperature at pH 7.5 of 100 deg. C. It has at
 CC least 90% activity after heat-treatment at 100 deg. C for 30 minutes at
 CC pH 7.5. It can transfer at least one glucose unit in alpha-1,4-glucan
 CC to alpha-1,4-glucan by an alpha-1,4-glucoside bond.

XX Sequence 653 AA;

alignment_scores:
 Quality: 121.50 Length: 386
 Ratio: 0.63 Gaps: 221
 Percent Similarity: 48.187 Percent Identity: 21.244

alignment_block:

US-09-886-400-3 x AAW54870 ..

Align seg 1/1 to: AAW54870 from: 1 to: 653

```

73 GAGAGGATACATCCAGCATCGAGACACTGATTAAAGAAATTC 122
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
28 GluArgSerTyrArgProPheMetGluThrLeu.....GluGlyTrp 42
123 T.....TTTGGCTCAACATAACGGCTATACCTTAAAGTTCCTC. 162
|  |||  |||  |||  |||  |||  |||  |||  |||  |||
42 OasnMetLysValAlaValHisTyrSerGlyProLeuLeuGluTrpIle 59
163 .....CCGAGGATATATAGACTCTCTTAAAGGGGCGCGAGT 204
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
59 rgAspAsnLysProGluHisLeuAspLeuLeuArgSerLeuValLysArg 75
205 GACCTGATAGATAATCGGAAGAGCTACACGACGACATCTCCCT 254
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
76 GlyGlnLeuGluIleValAlaGlyPheTyrGlyProValLeuAla 92
255 CTTCCGCTTAGCAGATAGACACAAAGTTCAGAGATAGGGAAGT 304
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
92 rIleProGluAspArgIleValGlnIleGluLysLeuLysGluPhe 109
305 AGGAAGAGCTCTCGAGCTTCTCCAAAGGATCTGGCTGCCAGAGCTC 354
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
109 laArgAsnLeu...GlyTyrGluAlaArgGlyValTrpLeuThrGluArg 124
355 GCCTATGACCGATAATCCCTGCCATCTACGAGGACACGCTTATGAGTA 404
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
125 ValTrpGlnProGluLeuValLysSerLeuArgAlaAlaGlyIleAsp 141
405 TCTATTCGCGACGGGAGCGGATGCTTTCTCAGCTCATCTCAACTCGG 454
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
141 rValIleValAsp.....AspTyrHisPheMetSerA 152
455 CG...ATAAGCCCAATTAACCGCTCTATCCACACCTTATAAGGCCAA 501
||  |||  |||  |||  |||  |||  |||  |||  |||  |||
152 laGlyLeuSerLysAspGluLeuPheTrpProTyrTrpThrGluAspGly 168
502 AGGNA.....AAGCGCTTAGGTACATCAG 527
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
169 GlyGluValIleThrValPheProIleAspGluLysLeuArg..... 182
528 CTATCTCTCTGCTCAGGAGGCTTAGGAAGCGGATAAAGCTCGTTT 576
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
183 TyrLeuIleProPheArgProValAspLysThrLeuGluTyrLeuHis 199
577 .....GAGGTAGGTAAAGTCAGTC.....AAAGCATC 612
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
199 erLeuAspAspGlyAspGluSerLysValAlaValPheHisAspAspGly 215
613 CAAGCGGTACCGTTTGGGTGGCGGTGAACACGCGCTGTAATGCTGGCAT 662
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

216 GluLysPheGlyValTrpProGlyThrTyrGluTrpValTyr..... 229
663 CGGAAGGCTTCTCTTATGAATCTAAGAAAGTGGGAGCTGGATA.... 708
230 .....GluLysGlyTrpLeuArg 236
709 .....GAGCACAGACACATCTTCTATAC 735
236 luPhePheAspArgValSerSerAspGluArgIleAsnLeuMetLeuTyr 252
736 GGCACCGATATAGAG.....TTTCATTGGCTATAGGACAT 770
253 SerGluTyrLeuGlnArgPheArgProArgGlyLeuValTyrLeuProI 269
771 TGCAGGCTAC...AGAATGAGTGTGAGGATTA..... 801
269 eAlaSerTyrPheGluMetSerGluTrpSerLeuProAlaArgGlnAla 286
802 .....TTAGAGTTATAGAGGAGCTCAACTCGGAAGTGTGCTTCCC 843
286 ysLeuPheValGluPheValGluLeuLys..... 296
844 TCAGAGCTGAAGCACAGTGGAGGAGCTCTACTTACGAGCTTCGAGTTG 893
297 LysGluAsnLysPheAspArgTyrArgValPheValArgGlyIleTr 313
894 GGCACACAGATAAGAGCTTGAGGATATGGAGAGGACGAGGGAAC... 939
313 p.....LysAsnPhe...PhePheLysTyrProGluSerAsnTyrM 326
940 .....GCAAGACTTATATCTCTCTACAAATATAGGGGCGGAGCTCGCC 984
326 eHisLysArgMetLeuMetValSerLysAlaValArgAsnAsn..... 340
985 CTTTTCGCGAGACACGATGCAAGGGGATGGAGCCCTCCCTCGAGAG 1034
341 .....ProGluAla 343
1035 GAGCTCGATGCTTCGGCGGATATATACAGAT.....TGGAGGGGTG 1078
343 aArgGluPheIleLeuArgAlaGlnCysAsnAspAlaTyrTrpHisGly 360
1079 AAAATGGG 1086
360 alPheGly 362
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.AAG82528
seq_documentation_block:
ID AAG82528 standard; Protein; 360 aa.
XX
AC AAG82528;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2150.
XX
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;
XX vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 990S-0164258.
XX (GLAX ) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;

```

XX WPI; 2001-316495/33.
DR N-PSDB; AAH53378.
XX
XX Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 18; Page 578-579; 2189pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG93120, from *Staphylococcus epidermidis*.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC *S. epidermidis* polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 360 AA;
SQ

```

148 .....GlyLeuAlaLysGluGlyAsnTyrGluT 157
404 ATCTATTTCGGCGAGGGAGCGCATGCTTTCTCACTCATCTCAACTCG 453
157 YrCySPhePheGluSerGluGly.....LysGly 166
454 CGCATAAAGCCAATTAAACCGCTCTCTCCACACCTATTAAAGGCCCAAG 503
157 GlnPheLysProValGlyAspAlaSerProTyrIleGlyLysLeuAspSe 183
504 GGAAGCGCTTAGGTAGTACATCAGTACTCTCTTGCTCTCAGGAGCTTA 553
183 r.....IleGluTyrVal..... 187
554 GGAAGCGGATAAAGCTCGTTT.....GAAGGTAAGGTA 588
188 ..AspGluIleLysLeuGluPheMetIleLysAspAsnGluLeuGluIle 203
589 ACGTAAGGCAAGTCAAGACATCGAA.....GCCGTACCCGTT..... 627
204 ThrLysArgAlaIleLeuAspAsnHisProTyrGluThrProValPheAs 220
628 .TGGTGGCGCTGAACAGCGCTCTAATGCTGGCATCGGAAGCTTCMC 676
220 pPheIleLysMetAsnLysGluSerGluTyrGlyLeuGlyIleIleGlyG 237
677 TTATGAATCTAAGAACTGGCAGCTGGATAGAGACAGACAGACACATT 726
237 InLeuAsnGlnThrMetThrLeuAspGluPheSerGluTyrAlaLysLys 253
727 CTTCTATAGCGCACCGCATATAGAGTTTCATTGGCTATAGGACATTGCAGG 776
254 GlnLeuAsnIleProSerValArgTyrThrGlyGlnHisAspSerProI 270
777 CTACAGATGAGTGTTCGAGGATATTAGAGTT.....A 811
270 eLysLysValAlaIleIleGlyGlySerGlyIleGlyPheGluTyrLysA 287
812 TAGACGAGCTCAACTCGGAAGTGTGCTTCCTCCACAGAGCTGAAGCAC 858
287 laSerGlnLeuGlyAlaAspValPheValThrGlyAspIleLysHis 302
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.AAG82506
seq_documentation_block:
XX
AC AAG82506;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:
XX
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
FI Kimberly WJ;
XX
DR WPI; 2001-316495/33.
XX
DR N-PSDB; AAH53356.
XX

```

PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
 PT useful for vaccinating against infections, e.g. endocarditis -
 XX
 PS Claim 18; Page 569; 2188pp; English.
 XX
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAH81454 to AAH83120, from *Staphylococcus epidermidis*.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC *S. epidermidis* polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 SQ Sequence 376 AA;

alignment_scores:
 Quality: 104.50 Length: 299
 Ratio: 0.731 Gaps: 12
 Percent Similarity: 47.826 Percent Identity: 19.398

alignment_block:
 US-09-886-400-3 x AAG82506 ..
 Align seg 1/1 to: AAG82506 from: 1 to: 376

```

10 CTCCTCTTTCACGCCACCTCCAGTATGCCGAATCCCAAGAGCGCAAT 59
|||||:|||||
47 LeuIlePheLysGly.....Va 52

60 CCCAAGGTCATAGAGGACATACATCCAGTCATCGACACATGATTA 109
: |||||:|||||
52 lLysArgileValGluAspGlyTyrglySerIleIleArgLysLeuIleG 69
|||||:|||||

110 AAGAGAAATCTTTTGGGCTCAACATACGGGCTATACCTTAAAGTTC 159
: |||||:|||||
69 lAsnAsnIle.....AsnLeuIleAlaLeuHisThrAsnLeuAspVal 83
|||||:|||||

160 CTCGCCAAGGATATTATAGACCTCGTTAAAGGGGCGATCGAGTCACCT 209
|||||:|||||
84 AsnProLysGlyValAsnArgMetLeuAlaAspGlnIleGlyLeuGluAs 100
|||||:|||||

210 GATGAGATATCCGAACG.....AGCTACAGCGACGCAATACTCCCC 253
|||||:|||||
100 nIleSerMetIleAsnThrAsnSerTyrTyrTyrLysValGlnThrP 117
|||||:|||||

254 TCCCTCCGCTTAGCAGAGTACAGACACAACTTCAGAGATCGGAATT 303
: |||||:|||||
117 heIleProLysAsnTyrlleGlu.....AspPhe 126
|||||:|||||

304 AAGGAAGAGCTCTTCGAGCTTTCTCCAAAGGATTTCTGGCTGCAGAGCT 353
|||||:|||||
127 LysAspSerLeuAsnGluLeu..... 133
|||||:|||||

354 GCGCTATGACCCGATATCCCTGCCATATGAGGACACAGGTTATGAGT 403
|||||:|||||
134 .....GlyLeuAlaLysGluGlyAsnTyrlleGlu 143
|||||:|||||

404 ATCTATTCCGCGACGGGAGCGAGTCTTTTCAGCTCATCAACTCG 453
|| ||| |||:|||||

```

```

143 yrcysPhePheGluSerGluGly.....LysGly 152
454 GCGATAAAGCAATTAACCGCTCTATCCACACCTTATAAGGCCCAAG 503
: |||||:|||||
153 GlnPheLysProValGlyAspAlaSerProTyrIleGlyLysLeuAspSe 169
|||||:|||||
504 GGAAGACCGCTTTAGGTACATCAGCTATCTCCTTGTCTCAGGAGCTTA 553
|||||:|||||
169 r.....IleGluTyrVal..... 173
554 GGAAGCGGATAAAGCTCGTWTTT.....GAAGTAAGGTA 588
|||||:|||||
174 ..AspGluIleLysLeuGluPheMetIleLysAspAsnGluLeuGluIle 189
|||||:|||||
589 ACGTAAAGCAGCTCAAGACATCGAA.....GCCGTACCCGTT..... 627
|||||:|||||
190 ThrLysArgAlaIleLeuAspAsnHisProTyrGluThrProValPheAs 206
|||||:|||||
628 .TGGGTGGCGCTGAACAGGCTGTATGTCTCGCATCGGAAGCTTCCTC 676
|||||:|||||
206 pPheIleLysMetAsnLysGluSerGluTyrGlyLeuGlyIleIleGlyG 223
|||||:|||||
677 TTATGAATCCTAAGAAAGTGGCAGCTGGATAGAGGACAAAGGACACATT 726
|||||:|||||
223 lLeuAsnGlnThrMetThrLeuAspGluPheSerGluTyrAlaLysLys 239
|||||:|||||
727 CTCCTATACGGCACCATATAGATTTTCATTTGGCTATAGGACATTCAGG 776
|||||:|||||
240 GlnLeuAsnIleProSerValArgTyrThrGlyGlnHisAspSerProIl 256
|||||:|||||
777 CTACAGATGAGTGTTCAGGAGATTATAGAGTT.....A 811
|||||:|||||
256 eLysLysValAlaIleIleGlyLysSerGlyIleGlyPheGluTyrLysA 273
|||||:|||||
812 TAGACGAGCTCAACTCGAACTGTGCTTCCTCCCTCAGAGCTGAAGCAC 858
|||||:|||||
273 laSerGlnLeuGlyAlaAspValPheValThrGlyAspIleLysHis 288
|||||:|||||

seq.name: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.AAR94013
seq_documentation_block:
ID AAR94013 standard; Protein; 329 AA.
XX
AC AAR94013;
XX
DT 21-AUG-1996 (first entry)
XX
DE Heat resistant maleate dehydrogenase.
XX
KW Heat resistant maleate dehydrogenase; h-rMAD; NADH; L-aspartic acid;
KW alpha-ketoglutaric acid; glutamine oxalo-transaminase activity; GOT.
XX
OS Bacillus stearothermophilus ATCC 12016.
XX
FH Key Location/Qualifiers
FT Misc-difference 185
FT /note= "Given in the specification as Var"
XX
PN JP08047389-A.
XX
PD 20-FEB-1996.
XX
PF 01-JUL-1994; 94JP-0151045.
XX
PR 03-JUN-1994; 94JP-0121629.
XX
PR 02-JUL-1993; 93JP-0164701.
XX
XX (TOYM ) TOYOBO KK.
XX
XX WPI; 1996-166248/17.
XX
XX N-PSDB; AAT17715.
XX
XX Protein having heat resistant maleate dehydrogenase activity - and
XX PT

```

PT	reagent contg. protein, NADH and L-aspartic and alpha-keto:glutaric acid for determination of glutamine oxalo-transaminase activity
PT	
XX	Claim 6; Page 13-15; 17pp; Japanese.
XX	
CC	This sequence represents a protein having heat resistant maleate dehydrogenase (h-rWAD) activity. The protein has a residual activity after storage at 40 deg.C for 10 days of at least 60%, pref. 70% and esp. 90%. A reagent containing the h-rWAD protein, NADH and L-aspartic and alpha-ketoglutaric acid may be used for the determination of glutamine oxalo-transaminase (GOT) activity. The h-rWAD protein may be produced by transforming E. coli with the DNA encoding this protein and isolating the protein from the culture medium.
CC	
XX	Sequence 329 AA;
XX	Sequence 329 AA;

```

alignment_scores:
  Quality: 104.00      Length: 154
  Ratio: 1.316        Gaps: 7
  Percent Similarity: 51.299  Percent Identity: 25.974

alignment_block:
  US-09-886-400-3 x AAR94013      ..

  Align seg 1/1 to: AAR94013 from: 1 to: 329

      28 CTC CAGATGTCGGAATATCCCAAGAGCGAATATCCCAAG ..... 66
      130 MetThrTyThrValPheLysGluSerGlyPheProLysAsnArgVal11 146
      146 eGlyGlnSerGlyValLeuAspThrAlaArg.....PheArgThrP 160
      167 .....GTCATAGAAGAGCATACATCCAGCATCTCGAGACAC 103
      146 eGlyGlnSerGlyValLeuAspThrAlaArg.....PheArgThrP 160
      104 TGATTAAAGAAAGAAATTCCTTTGGGCTCAACATAACGGGCTATACCTTA 153
      160 heValAlaGluGlu.....LeuAsnIleSer..... 168
      154 AAGTTCCTCCGAGGATATATATAGACTCGTTAAAGGGGGCATCCCGAG 203
      169 .....ValLysAspValThrGlyPheValLeuGlyGlyHisGlyAs 182
      204 TGACCTGATAGAGATATCGGAAGCAGCTACACGCACGCAATA.....C 247
      182 pAspMetValProLeuValArgTyrSerTyrAlaGlyGlyIleProLeuG 199
      248 TCCCCCTCCTCCCGCTTAGCAGATAGACGACAGATTCAGAGAGATAGG 297
      199 LuLysLeuIleProLysAspArgLeuAspAlaIleValGluArgThrArg 215
      298 GAAGTTAAGGAAGAGCTTCGAGCTTCTCCAAAGGGA.....TTCG 341
      216 LysGlyGlyGlyGluIleValAsnLeuLeuGlyAsnGlySerAlaTyr 232
      342 GCTGCCAGAGCTGCCTATGACCCGATATCCCTCGCTACTCAGGAGCA 391
      232 rAlaProAlaAlaSerLeuValGluMetValGluAlaIleLeuLysAsp 249
      392 AC.....GCTTAT 399
      249 lnArgArgIleLeuProAlaIleAlaTyrLeuGluGlyGluTyrGly 265
      400 GATGATCTATTC 411
      256 GluGlyIleTyr 269

```

```
seq_name: /SIBS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA616038
seq_documentation_block:
ID   AAG16038 standard; Protein; 571 AA.
XX
AC   AAG16038;
```

XX	17-OCT-2000	(first entry)
DT XX		
DE XX	Arabidopsis thaliana protein	
EX XX	protein identification; sign	
KW KW	hybridisation assay; genetic	
KW KW	termination sequence.	
XX		
OS OS	Arabidopsis thaliana.	
XX XX		
PX PN	EPI033405-A2.	
PD PD	06-SEP-2000.	
XX XX	25-FEB-2000;	2000EP-0301439.
PF PF	05-MAR-1999;	99US-0121825.
PR PR	05-MAR-1999;	99US-0123180.
PR PR	09-MAR-1999;	99US-0123544.
PR PR	23-MAR-1999;	99US-0125788.
PR PR	25-MAR-1999;	99US-0126264.
PR PR	29-MAR-1999;	99US-0126785.
PR PR	01-APR-1999;	99US-0127462.
PR PR	06-APR-1999;	99US-0128234.
PR PR	08-APR-1999;	99US-0128714.
PR PR	16-APR-1999;	99US-0129845.
PR PR	19-APR-1999;	99US-0130077.
PR PR	21-APR-1999;	99US-0130449.
PR PR	23-APR-1999;	99US-0130510.
PR PR	23-APR-1999;	99US-0130891.
PR PR	28-APR-1999;	99US-0131449.
PR PR	30-APR-1999;	99US-0132048.
PR PR	30-APR-1999;	99US-0132407.
PR PR	04-MAY-1999;	99US-0132484.
PR PR	05-MAY-1999;	99US-0132485.
PR PR	06-MAY-1999;	99US-0132486.
PR PR	06-MAY-1999;	99US-0132487.
PR PR	11-MAY-1999;	99US-0132863.
PR PR	11-MAY-1999;	99US-0134256.
PR PR	14-MAY-1999;	99US-0134218.
PR PR	14-MAY-1999;	99US-0134219.
PR PR	14-MAY-1999;	99US-0134221.
PR PR	14-MAY-1999;	99US-0134370.
PR PR	18-MAY-1999;	99US-0134768.
PR PR	20-MAY-1999;	99US-0134941.
PR PR	20-MAY-1999;	99US-0135124.
PR PR	21-MAY-1999;	99US-0135353.
PR PR	24-MAY-1999;	99US-0135629.
PR PR	25-MAY-1999;	99US-0136021.
PR PR	27-MAY-1999;	99US-0136392.
PR PR	28-MAY-1999;	99US-0136782.
PR PR	01-JUN-1999;	99US-0137222.
PR PR	03-JUN-1999;	99US-0137528.
PR PR	04-JUN-1999;	99US-0137502.
PR PR	07-JUN-1999;	99US-0137724.
PR PR	08-JUN-1999;	99US-0136094.
PR PR	10-JUN-1999;	99US-0138540.
PR PR	10-JUN-1999;	99US-0138687.
PR PR	14-JUN-1999;	99US-0139119.
PR PR	16-JUN-1999;	99US-0139452.
PR PR	16-JUN-1999;	99US-0139453.
PR PR	17-JUN-1999;	99US-0139454.
PR PR	18-JUN-1999;	99US-0139455.
PR PR	18-JUN-1999;	99US-0139457.
PR PR	18-JUN-1999;	99US-0139458.
PR PR	18-JUN-1999;	99US-0139459.
PR PR	18-JUN-1999;	99US-0139460.
PR PR	18-JUN-1999;	99US-0139461.
PR PR	18-JUN-1999;	99US-0139462.
PR PR	18-JUN-1999;	99US-0139463.
PR PR	18-JUN-1999;	99US-0139464.

fragment SEQ ID NO: 16524.
transduction pathway; metabolic pathway;
mapping; gene expression control; promoter


```
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145215.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146385.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 104.00 Length: 365
Ratio: 0.638 caps: 20
Percent Similarity: 44.658 Percent Identity: 22.466

alignment_block:
US-09-886-400-3 x AAG16038 ..
Align seg 1/1 to: AAG16038 from: 1 to: 571

85 ATCCAGTCATCGACACTGATTAAGAAGAAATCTTTTGGGCTCAA 134
101 LeuProileLeuGlu...MetValProGluGluLeuHisPhe..... 113
135 CATACACGGGCTATACCTAAAGTTCTCCCGAAGGATATTATAGACCTCG 184
114 ThrValProTyrAlaLeuSerPheileMetGlyAspProileLysMetA 130
185 TTAAGGGGCGATCGCGAGTGACCTG.....ATAGAGATA 219
130 lathrLeuGlyIleAspAsnGlnLeuProThrGlyVallysileGlyLys 146
```

220 ATCGGAACGAGCTACAGCAGCAATACTACTCCCTCTCCCGCTTAGCAG 269
::: ||| :::::::::::::::::::: |
147 LeuArgGlnAlaGLeuthrLysThrMetLeuProLeu.....LeuSerGI 161

270 AGTAGAACACAAGTTCCAGAGAGT..... 294
::: ::: |||||
161 uLeuGlyGlyLeileProArgGluThrLeuLeuTrpLysLeuLysLeuL 178

295AGGGAAGTTAAGAA 309
::: |||||
178 euArgSerGlyCysAlaTrpAlaAsnSerArgIleHisAlaValGlnAla 194

310 GAGCTTCCTCGAGCTT...TCTCCAAGGGATTCTGCCTGCCAGAGCTCGC 356
||||::: ||| ||| ::| |||||:::
195 GluValLeuValLeuAlaSerGlyLysAspMetLeuProSerGlnGl 211

357 CTATGACC CGATATCCCTCGCATCTG..... 384
::: |||||
211 uGluAlaLysArgLeuHisGlyLeuLeuLysAsnCysSerValArgCysP 228

385 ..AAGACAACGGTTATGAGTATCTATTCGCCGACGGGAGCGCATGCTT 432
||||::: |||||::: |||||::: |||||
228 helyAspAsnGlyHisThrLeuLeuLeuGluAspSerIleSerLeuLeu 244

433 TTCTCAGCTCATCTCAACTCGGCATATAACCCAAATTAACCGCTCTATCC 482
::: |||||
245ThrValIleGlyGlyThrGly..... 251

483 ACACCTTATAAGGCCCAAGAAAAGCGCTTTAGGTACATCAGC.... 528
||| :::::::::::::::::::: |
252LysTy rArgArgSerTrpArgTy rAspLeuValSerAsp 265

529TATCTCTTGCT.....CTC 543
||| |||||
265 heLeuProSerLysGlyGluLeuAlaTrpAlaLeuAspGluValLeu 281

544 AGGAGCTTTAGGAAGCGCATAAAGCTCGTTTT.....GA 578
||||::: |||||::: |||||
282 GlyPheLeuArgAsnAlaValGlySerValPhePheSerThrMetGluAs 298

579 AGTAGAGTTAAGCTTAAGGCAGTCAAAGCACATCAAGCCGTCACCC.... 524
||||::: |||||::: |||||
298 pGlyLysIle.....ValLysGlyLeuAlaGlyValProAspL 311

625GTTTGGTGCGCTGAACACGGCTGTATGCTCGGCATCGGAAGG 669
|||::: |||||::: |||||
311 ysglyProValLeuLeuValGlyTy rHisMetLeuMetGlyLeuGluLeu 327

670 CTTCTCTTATGAATCCTAAGAAGTGGCGAGCTGGATAGAGCAAGA 719
||||::: |||||
328 GlyProMetSerGlu.....AlaPheIleLysGluLys.. 338

720 CAACATCTCTCA..... 732
|||||
339 .AsnIleLeuPheArgGlyMetAlaHisProValLeuTy rSerAspAsnA 355

733TACGGCACCGATATAGATTTCATTGGCTAT 762
||||| |||||
355 spProAlaLysAlaPheAspTy rGlyAspTrpIleLysValPheGly... 370

763 AGGACATTCCAGGCTACAGATGAGTCTTCAGGATATATAGAGTTAT 812
||||| |||||
371AlatyrProValThrAlaThrAsnLeuPheLysLeule 383

813 AGACAGCTCAACTCGGAACCTTGCTTCCCTCA.....GAGCTGA 853
||||::: ||| ||| |||||
383 uAspSerLysSerHisValLeuLeuPheProGlyGlyAlaArgGluAlal 400

854 AGCACAGTGAAGGAGGAGCTCTACTTACCGAGCTTCGAGTTGGGCACCAGT 903
||||::: ||| |||
400 euHisAsnArgglyGluGlnTy r..... 407

```

904 AAGACCTTGAGGATATGAGAGCAGGACGACGGACCGAAGACTT 948
      ||||| |||:::~::~|||||:::
408 .....LysLeulleirProGlugInglInluPheValArgMet 420

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AA48655
seq_documentation_block:
ID   AG48655 standard; Protein; 571 AA.
XX   AC   AAG48655;
XX
XX   DT   18-OCT-2000 (first entry)
XX
XX   DE   Arabidopsis thaliana protein fragment SEQ ID NO: 61465.
XX
XX   KW   Protein identification; signal transduction pathway; metabolic pathway;
KW       hybridisation assay; genetic mapping; gene expression control; promoter;
KW       termination sequence.
XX
XX   OS   Arabidopsis thaliana.
XX
XX   PN   EP1033405-A2.
XX
XX   PD   06-SEP-2000.
XX
XX   PE   25-FEB-2000; 2000EP-0301439.
XX
XX   PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   28-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132487.
PR   07-MAY-1999; 99US-0132863.
PR   11-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   27-MAY-1999; 99US-0136392.
PR   28-MAY-1999; 99US-0136782.
PR   01-JUN-1999; 99US-0137222.
PR   03-JUN-1999; 99US-0137526.
PR   04-JUN-1999; 99US-0137502.
PR   07-JUN-1999; 99US-0137724.
PR   08-JUN-1999; 99US-0138094.
PR   10-JUN-1999; 99US-0138540.
PR   10-JUN-1999; 99US-0138847.
PR   14-JUN-1999; 99US-0139119.
PR   16-JUN-1999; 99US-0139452.
PR   16-JUN-1999; 99US-0139453.
PR   17-JUN-1999; 99US-0139492.

```

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139818.
PR 21-JUN-1999; 99US-0139819.
PR 21-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 25-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:

Quality: 104.00

Length: 365

Ratio: 0.638

Gaps: 20

Percent Similarity: 44.658 Percent Identity: 22.466

alignment_block:

US-09-886-400-3 x AAG48655 ..

Align seg 1/1 to: AAG48655 from: 1 to: 571

85 ATCCGAGTCATCGAGACACTGATTAAAGAGAATAATCCCTTTGGGTCAA 134

```

101 LeuProIleLeuGlu...MetValProGluGluLeuHisPhe..... 113
135 CATACGGGCTATACCTTAAGATTCTCCCGAAGATATATACACTCG 184
114 ThrValProTyrAlaLeuSerPheIleMetGlyAspProIleLysMet 130
185 TTAAGGGGGCATCCGAGTACACCTG.....ATAGAGATA 219
130 IatThrLeuGlyIleAspAsnGlnLeuProThrGlyValLysIleGluLys 146
220 ATCGGAACAGAGTACACGACGCAATATCTCCCTCTCCCGCTAGCAG 269
147 LeuArgGlnArgLeuThrLysThrMetLeuProLeu.....LeuSerGI 161
270 AGTAGAGCACACAGTTCAGAGAGAT..... 294
161 uLeuGlyGlyIleIleProArgGluThrLeuLeuTrpLysLeuLysLeu 178
295 .....AGGGAAGTTAAGAA 309
178 euArgSerGlyCysAlaTyrAlaAsnSerArgIleHisAlaValGlnAla 194
310 GAGCTCTTCGAGCTT...TCTCCAAGGATTCGCTGCCGACAGCTCGC 356
195 GluValLeuValLeuAlaSerGlyLysAspMetMetLeuProSerGlnG 211
357 CTATGACCGGATAATCCCTGCCATACTG..... 384
211 uGluAlaLysArgLeuHisGlyLeuLeuLysAsnCysSerValArgCysp 228
385 ..AAGGACACAGGTTATGAGTATCATTCGCGGAGGGGAGCGCATGCTT 432
228 helyAspAsnGlyHisThrLeuLeuLeuGluAspSerIleSerLeuLeu 244
433 TTCTCAGCTCATCTCACTCGCGGATAAAGCAATTAACCGCTCTATCC 482
245 .....ThrValIleLysGlyThrGly..... 251
483 ACACCTTATAAGGCCCAAGGAAACGCTTTAGGTACATCAGC..... 528
252 .....LysTyrArgSerTrpArgTyrAspLeuValSerAsp 265
529 .....TATCTCCTTGGT.....CTC 543
265 heLeuProProSerLysGlyGluLeuAlaTyrAlaLeuAspGluValLeu 281
544 AGGAGCTTAGGAGCGGATAAAGCTCGTTT.....GA 578
282 GlyPheLeuArgAsnAlaValGlySerValPhePheSerThrMetGluAs 298
579 AGTAAAGTAAAGCTAAAGGCGATCAAGACATCAAGCCCTACCC..... 624
298 pGlyLysIle.....ValLysGlyLeuAlaGlyValProAspL 311
625 .....GTTTGGTGGCGCGTGAACAGCGTGTAAATCCTCGGATCGGAAG 669
311 yselProValLeuLeuValGlyTyrHisMetLeuMetGlyLeuGluLeu 327
670 CTTCTCTTATGATCCTTAAGAAAGTGGCGAGCTGGATAGAGGACAGGA 719
328 GlyProMetSerGlu.....AlaPheIleLysGluLys... 338
720 CAACATCTCTCTA..... 732
339 AsnIleLeuPheArgGlyMetAlaHisProValLeuTyrSerAspAsnA 355
733 .....TACGGCAGGATATAGATTCTATTCGCTAT 762
355 spProAlaLysAlaPheAspTyrGlyAspTrpIleLysValPheGly... 370
763 AGGACATTCACGGCTACAGATGACTGTTGAGGATATTAGAGTTAT 812

```

```

371 .....AlaTyrProValThrAlaThrAsnLeuPheLysLeuLe 383
813 AGACGAGCTCAACTCGGAACTGCTCCCTCCCTCA.....GAGCTGA 853
383 uaspSerLysSerHisValLeuLeuPheProGlyLysAlaArgGluAlaL 400
854 AGCACAGTGGAGGAGCTCTACTTACGAGCTTCAGTTGGGCACAGAT 903
400 euHisAsnArgGlyGluGlnTyr..... 407
904 AAGACCTTGAGATATGGAGAGAGGACGACCAAGGACGACGACTT 948
408 .....LysLeuIleTrpProGluGlnGlnGluPheValArgMet 420
seq_name: /SIBS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA616037
seq_documentation_block:
ID: AA616037 standard; Protein; 704 AA.
XX
AC: AAG16037;
XX
DT: 17-OCT-2000 (first entry)
XX
DE: Arabidopsis thaliana protein fragment SEQ ID NO: 16523.
XX
KW: Protein identification; signal transduction pathway; metabolic pathway;
KW: hybridisation assay; genetic mapping; gene expression control; promoter;
KW: termination sequence.
XX
OS: Arabidopsis thaliana.
XX
PN: EPI033405-A2.
XX
PD: 06-SEP-2000.
XX
PF: 25-FEB-2000; 2000EP-0301439.
XX
PR: 25-FEB-1999; 99US-0121825.
PR: 05-MAR-1999; 99US-0123180.
PR: 09-MAR-1999; 99US-0123548.
PR: 23-MAR-1999; 99US-0125788.
PR: 25-MAR-1999; 99US-0126264.
PR: 29-MAR-1999; 99US-0126785.
PR: 01-APR-1999; 99US-0127462.
PR: 06-APR-1999; 99US-0128234.
PR: 08-APR-1999; 99US-0128714.
PR: 16-APR-1999; 99US-0129845.
PR: 19-APR-1999; 99US-0130077.
PR: 21-APR-1999; 99US-0130449.
PR: 23-APR-1999; 99US-0130510.
PR: 28-APR-1999; 99US-0130891.
PR: 30-APR-1999; 99US-0131449.
PR: 30-APR-1999; 99US-0132048.
PR: 30-APR-1999; 99US-0132407.
PR: 04-MAY-1999; 99US-0132484.
PR: 06-MAY-1999; 99US-0132485.
PR: 06-MAY-1999; 99US-0132486.
PR: 07-MAY-1999; 99US-0132863.
PR: 11-MAY-1999; 99US-0134256.
PR: 14-MAY-1999; 99US-0134218.
PR: 14-MAY-1999; 99US-0134219.
PR: 14-MAY-1999; 99US-0134221.
PR: 14-MAY-1999; 99US-0134370.
PR: 18-MAY-1999; 99US-0134768.
PR: 19-MAY-1999; 99US-0134941.
PR: 20-MAY-1999; 99US-0135124.
PR: 21-MAY-1999; 99US-0135353.
PR: 24-MAY-1999; 99US-0135629.
PR: 25-MAY-1999; 99US-0136021.
PR: 27-MAY-1999; 99US-0136392.
PR: 28-MAY-1999; 99US-0136782.
PR: 01-JUN-1999; 99US-0137222.

```

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452..
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145216.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156595.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160988.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:

Quality: 104.00 Length: 365
Ratio: 0.538 Gaps: 20
Percent Similarity: 44.658 Percent Identity: 22.466

alignment_block:

US-09-886-400-3 x AAG16037

Align seg 1/1 to: AAG16037 from: 1 to: 704

```
85 ATCCAGTCATCGACACACTGATTAAAGCAAAATTCCTTTGGGCTCAA 134
234 LeuProIleLeuGlu...MetValProGluGluLeuHisPhe..... 246
135 CATAACGGGCTATACCTAAAGTCTCTCCCGAAGGATATATAGACCTCG 184
247 ThrValProTyrAlaLeuSerPheIleMetGlyAspProIleLysMetA 263
185 TTAAGGGGGCATCCGGAGTACCTG.....ATAGAGATA 219
263 IatThrLeuGlyIleAspAsnGlnLeuProThrGlyValLysIleGluLys 279
220 ATCGGAACAGCTACACGACGACCAATACCTCCCTCCCGCTAGCAG 269
280 LeuArgGlnArgLeuThrLysThrMetLeuProLeu.....LeuSerGI 294
270 AGTAGAAGCAACAGTTCAGAGAGAT..... 294
294 uLeuGlyGlyIleIleProArgGluThrLeuLeuThrLysLeuLysLeu 311
295 .....AGGGAAGTTAAGGAA 309
311 euArgSerGlyCysAlaTyrAlaAsnSerArgIleHisAlaValGlnAla 327
310 GAGCTCTTCGAGCTT...TCTCCAAGAGGATCTGTGGTCCGAGAGCTGCG 356
328 GluValLeuValLeuAlaSerGlyLysAspMetMetLeuProSerGlnGI 344
357 CTATGACCCGATAAATCCCTGCCATCTG..... 384
344 uGluAlaLysArgLeuHisGlyLeuLeuLysAsnCysSerValArgCysP 361
385 ..AAGGACACGCTTATGATGATCTATTCCGCGACGGGAGCGGATGCTT 432
361 heLysAspAsnGlyHisThrLeuLeuLeuGluAspSerIleSerLeuLeu 377
433 TTCTCAGCTCATCTCACTCGCGGATAAAGCAATTAACCGCTCTATCC 482
378 .....ThrValIleLysGlyThrGly..... 384
483 ACACCTTTATAAGGCCCAAGGAAAGCGCTTTAGGTACATCAGC.... 528
385 .....LysTyrArgArgSerTrpArgTyrAspLeuValSerAspP 398
529 .....TATCTCCTTGGT.....CTC 543
398 heLeuProProSerLysGlyGluLeuAlaTyrAlaLeuAspGluValLeu 414
544 AGGAGCTTAGGAAGCGGATAAAGCTCGTTT.....GA 578
415 GlyPheLeuArgAsnAlaValGlySerValPhePheSerThrMetGluAs 431
579 AGGTAAAGTAAACGCTAAAGCAGCTCAAGACATCGAAGCCGTACCC.... 624
431 pGlyLysIle.....ValLysGlyLeuAlaGlyValProAspL 444
625 .....GTTTCGGTGGCGGTGAACACCGCTGTAATCGCTCGCATCGGAAGG 669
444 ysGlyProValLeuValGlyTyrHisMetLeuMetGlyLeuLeu 460
670 CTTCTCTTATGAATCTTAGAAGTGGCGAGCTGGATAGAGACAAGGA 719
461 GlyProMetSerGlu.....AlaPheIleLysGluLys... 471
```

```
720 CAACATTCTTCTA..... 732
|||||
472 AsnIleLeuPheArgGlyMetAlaHisProValLeuTyrSerAspAsnA 488
733 .....TACGCGACCGATATAGAGTTTCATTGGCTAT 762
488 spProAlaLysAlaPheAspTyrGlyAspTrpIleLysValPheGly... 503
763 AGGACATTGCGAGCTACAGAAATGAGTGTGAGGAGTATTAGAGGTAT 812
504 .....AlaTyrProValThrAlaThrAsnLeuPheLysLeuLe 516
813 AGACAGAGCTCACTCGGAACGTGCTCCCTCCCTCA.....GAGCTGA 853
516 uaspSerLysSerHisValLeuLeuPheProGlyGlyAlaArgGluAlaL 533
854 AGCAGAGTGGGAAGGAGGAGCTACTTACGAGTTCGAGTTGGGCACAGAT 903
533 euHisAsnArgGlyGluGlnTyr..... 540
904 AAGAGCTTGAGGATATGGAGAGAGGACGAGGACGACGACACTT 948
541 .....LysLeuIleTrpProGluGlnGlnGluPheValArgMet 553
seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: AAG48654
seq_documentation_block:
ID AAG48654 standard; Protein: 704 AA.
XX
AC AAG48654;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61464.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
```

PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147132.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148568.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

```

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
  Quality: 104.00      Length: 365
  Ratio: 0.638        Gaps: 20
  Percent Similarity: 44.658  Percent Identity: 22.466

alignment_block:
US-09-886-400-3 x AAG15036  ..
Align seg 1/1 to: AAG15036 from: 1 to: 726

85 ATCCAGTCATCGACACTGNTAAAGAAATTCCTTTGGGCTCA 134
  :::::::::::::::::::: |||
256 LeuProIleLeuGlu...MetValProGluGluLeuHisPhe..... 268
135 CATAAGGGCTATACCTTAAAGTCTCCCGAAGGATATTATAGACPCG 184
  :::::::::::::::::::: |||
269 ThrValProTyrAlaLeuSerPheIleMetGlyAspProIleLysMetA 285
185 TTAAGGGGTCGCGAGTACGCTG.....ATAGAGATA 219
  :::::::::::::::::::: |||
285 IaThrLeuGlyIleAspAsnGlnLeuProThrGlyValLysIleGluLys 301
220 ATCGGAACGAGTCACGACGACATACCTCCCTCCGCTAGCAG 269
  :::::::::::::::::::: |||
302 LeuArgGlnArgLeuThrLysThrMetLeuProLeu.....LeuSerGI 316
270 ACTAGAAGCACAAAGTTCAGAGAGAT..... 294
316 uLeuGlyGlyIleIleProArgGluThrLeuLeuTrpLysLeuLysLeuL 333
295 .....AGGAAAGTTAAGGAA 309
333 euArgSerGlyCysAlaTyrAlaAsnSerArgIleHisAlaValGlnAla 349
310 GAGCTCTTCGAGCTT...TCTCCAAAGGATTCGCTGCCAGAGCTCGC 356
  :::::::::::::::::::: |||
350 GluValLeuValLeuAlaSerGlyLysAspMetMetLeuProSerGlnGI 366
357 CTATGACCGGATAATCCCTGCCACTG..... 384
366 uGluAlaLysArgLeuHisGlyLeuLysAsnCysSerValArgCysP 383
385 ..AAGGACACGCTTATAGTATCTATTCGCGACGGGAGCGGATGCTT 432
  :::::::::::::::::::: |||
383 helysAspAsnGlyHisThrLeuLeuGluAspSerIleSerLeuLeu 399
433 TTCTCAGCTCATCTCAACTCGGCGATAAGCAATTAACCGCTCTATCC 482
  :::::::::::::::::::: |||
400 .....ThrValIleLysGlyThrGly..... 406
483 ACACCTTATAAGGCCCAAGGAAAGCGCTTAGGTACATCAGC..... 528
407 .....LysTyrArgArgSerTrpArgTyrArgIleValSerAsp 420

```

```

529 .....TATCTCCTTGGT.....CTC 543
420 heLeuProProSerLysGlyGluLeuAlaTyrAlaLeuAspGluValLeu 436
544 AGGGAGCTTAGAAGCGCATAAAGCTCGTTTT.....GA 578
437 GlyPheLeuArgAsnAlaValGlySerValPheSerThrMetGluAs 453
579 AGGTAAGGTAAAGCTAAAGGAGTCAAAAGCATCGAACCGGTACCC.... 624
  :::::::::::::::::::: |||
453 polyLysile.....ValLysGlyLeuAlaGlyValProAspL 466
625 .....GTTGGGTGCCCTGAACACGGCTGTAATGCTCGGCATCGGAAG 669
466 ysglyProValLeuLeuValGlyTyrHisMetLeuMetGlyLeuGluLeu 482
670 CTCTCTTATGAATCTTAAGAAAGTGGCGAGCTGGATAGAGGACAAGGA 719
  :::::::::::::::::::: |||
483 GlyProMetSerGlu.....AlaPheIleLysGluLys... 493
720 CAACATTCTTCTA..... 732
494 AsnIleLeuPheArgGlyMetAlaHisProValLeuTyrSerAspAsnA 510
733 .....TACGGCACCGATATAGAGTTTCATTTGGCTAT 762
510 spProAlaLysAlaPheAspTyrGlyAspTrpIleLysValPheGly... 525
763 AGGGACATTGCGAGCTACAGATGAGTGTGAGGATATTAGAGTTAT 812
526 .....AlaTyrProValThrAlaThrAsnLeuPheLysLeuLe 538
813 AGACGAGCTCAACTCGGAACGTGCTTCCCTCA.....GAGCTGA 853
538 unaspSerLysSerHisValLeuLeuPheProGlyGlyAlaArgGluAlaL 555
854 ACACAGTGGAAAGGAGCTCTACTACGAGCTTCGAGTTGGCACCAGAT 903
555 euHisAsnArgGlyGluGlnTyr..... 562
904 AAGAGCTTGAGGATATGGAGAGACGACGAAGGAACGCAAGACTT 948
563 .....LysLeuIleTrpProGluGlnGlnGluPheValArgMet 575

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: AAG48653
seq_documentation_block:
ID AAG48653 standard; Protein; 726 AA.
XX
AC AAG48653;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61463.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.

```

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 15-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141642.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

```
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

alignment_scores:
Quality: 104.00 Length: 365
Ratio: 0.638 Gaps: 20
Percent Similarity: 44.658 Percent Identity: 22.466

alignment_block:
US-09-886-400-3 x AAG48653
Align seg 1/1 to: AAG48653 from: 1 to: 726

```
85 ATCCAGTCATCGACACATGATTAAAGAAATCTCTTTGGGTCAA 134
256 LeuProLleLeuGlu...MetValProGluGluLeuHisPhe..... 268
135 CATACGGGGTATACCTTAAGTTCTCCGAGAGATATTATAGACTCG 184
269 ThrValProFyrAlaLeuSerPheIleMetGlyAspProIleLysMeta 285
185 TTAAGGGGCGATCGCGAGTGACCTG.....ATAGAGATA 219
285 laThrLeuGlylleAspAsnGlnLeuProThrGlyValLysIleGluLys 301
220 ATCGAGACGAGCTACACGCGCAATACCTCCCTCCCTCCGCTTAGCAG 269
302 LeuArgGlnArgLeuThrLysThrMetLeuProLeu.....LeuSerGI 316
270 AGTAGAACCAAGTTCAGACAGAT..... 294
316 uLeuGlylleIleProArgGluThrLeuLeuThrLysLeuLysLeuL 333
295 .....AGGGAAGTTAAGGAA 309
333 euArgSerGlyCysAlaTyrAlaAsnSerArgIleHisAlaValGlnAla 349
310 GAGCTCTTCGAGCTT...TCTCCAAAGGATTTGGCTGCCAGAGCTCGC 356
350 GluValLeuValLeuAlaSerGlyLysAspMetMetLeuProSerGlnGI 366
357 CTATGACCCGATATCCCTGCCATACG..... 384
366 uGluAlaLysArgLeuHisGlyLeuLeuLysAsnCysSerValArgCysP 383
385 ..AAGGACACGGTTATGAGTATCTATTCGCCGACGGGAGGCGATCGTT 432
```

```
383 heLysAspAsnGlyHisThrLeuLeuLeuGluAspSerIleSerLeuLeu 399
433 TTCAGCTCATCTCAACTCGCGATATAAGCCCAATTAAACCGCTCTATCC 482
400 .....ThrValIleLysGlyThrGly..... 406
483 ACACCTTATAAAGCCCAAGGAAAGCCCTTTAGCTACATCAGC..... 528
407 .....LysTyrArgArgSerTrpArgTyrAspLeuValSerAsp 420
529 .....TATCTCCTTGGT.....CTC 543
420 heLeuProProSerLysGlyGluLeuAlaTyrAlaLeuAspGluValLeu 436
544 AGGAGCTTAGAAGCGGATAAAGCTCGTTTT.....GA 578
437 GlyPheLeuArgAsnAlaValGlySerValPhePheSerThrMetGluAs 453
579 AGGTAAGCTTAAGCTAAAGGACGATCAAGACATCGAAGCCGTACCC..... 624
453 pGlyLysIle.....ValLysGlyLeuAlaGlyValProAspL 466
625 .....GTTTGGTGGCGGTGAACACAGCGCTGTAATGCTCGGCATCGGAAG 669
466 ysGlyProValLeuLeuValGlyTyrHisMetLeuMetGlyLeuGluLeu 482
670 CTTCCTCTTATGATCTCAAGAAAGTGGCGAGCTGGATAGAGGACAGGA 719
483 GlyPrometSerGlu.....AlaPheIleLysGluLys.. 493
720 CAACATCTTCTTA..... 732
494 AsnIleLeuPheArgGlyMetAlaHisProValLeuTyrSerAspAsnA 510
733 .....TACGGCACCATATAGAGTTCATTCGCTAT 762
510 spProAlaLysAlaPheAspTyrGlyAspTrpIleLysValPheGly... 525
763 AGGACATTGCAGGCTACAGAAATGAGTGTGAGGATATTAGAGGTAT 812
526 .....AlaTyrProValThrAlaThrAsnLeuPheLysLeuLe 538
813 AGACAGCTCAACTCGGACACTGTGCTCCCTCCCTCA.....GAGCTGA 853
538 uAspSerLysSerHisValLeuLeuPheProGlyGlyAlaArgGluAlaL 555
854 AGCAGGTGGAAGGAGCTCTACTTACGAGCTTCGAGTTCGGCACCAGAT 903
555 euHisAsnArgGlyGluGlnTyr..... 562
904 AAGAGCTTAGGATATGAGAGAGACGACGAAGGAAACGCAAGACTT 948
563 .....LysLeuIleTrpProGluGlnGlnGluPheValArgMet 575
```

seq_name: /SDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG18270
seq_documentation_block:
ID ABG18270 standard; Protein; 936 AA.
XX
AC ABG18270;
XX
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #18261.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.

OM of: US-09-886-400-3 to: SPTREMBL19:* out_format : pfs

Date: Jun 11, 2002 10:39 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet_n2p.model -DEV=xlh
-O/Cgn2.1/USPTC.spool/US09886400/tunat.11062002.150916.17416/app_query.fasta_1.1169
-DB-SPTREMBL19 -OEMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINWATCH=0.100 -LOOCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09886400.cgn1.1.219 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT_THREADS=1

Search information block:

Query: US-09-886-400-3

Query length: 1095

Database: SPTREMBL19.*

Database sequences: 562222

Database length: 172994529

Search time (sec): 61.230000

score_list:	Strid Orig	ZScore	EScore	Len	Documentation
Sequence					
sp_archaea:Q9HHB5	+ 1532.00	2383.59	5.6e-125	364	! Q9HHB5 pyrococcus furiosus. alp
sp_bacteriap:O58106	+ 1494.00	2324.17	1.1e-121	364	! O58106 pyrococcus horikoshii. h
sp_bacteriap:P74630	+ 162.50	238.26	1.2e-05	529	! P74630 synecocystis sp. (strai
sp_bacteriap:Q97BM4	+ 149.50	221.18	0.0001	378	! Q97BM4 thermoplasma volcanium.
sp_bacteriap:Q97GF3	+ 147.00	214.06	0.0003	527	! Q97GF3 clostridium acetobutylic
sp_bacteriap:Q973T0	+ 142.50	208.70	0.0006	443	! Q973T0 sulfolobus tokodaii. hyp
sp_bacteriap:Q972N0	+ 136.50	192.52	0.0025	895	! Q972N0 sulfolobus tokodaii. hyp
sp_bacteriap:Q50094	+ 130.50	187.67	0.0074	560	! O50094 pyrococcus horikoshii. h
sp_bacteriap:Q9K004	+ 128.00	178.92	0.0139	923	! Q9K004 bacillus halodurans. bhl
sp_fungi:Q9C105	+ 127.50	175.32	0.0164	1236	! Q9C105 schizosaccharomyces pom
sp_bacteriap:Q9V294	+ 124.00	168.91	0.0340	1362	! Q9V294 pyrococcus abyssi. anyl
sp_bacteriap:Q9HL91	+ 122.50	179.51	0.0332	357	! Q9HL91 thermoplasma acidophilum
sp_bacteriap:Q30246	+ 120.50	177.32	0.0485	324	! Q30246 archaeoglobus fulgidus.
sp_bacteriap:Q9V0M7	+ 119.00	168.98	0.0760	602	! Q9V0M7 pyrococcus abyssi. hypot
sp_bacteriap:Q93377	+ 111.00	157.78	0.3663	526	! Q93377 treponema pallidum. cons
sp_bacteriap:Q972D2	+ 109.50	157.01	0.4759	447	! Q972D2 sulfolobus solfataricus.
sp_invertebrate:Q9BKV7	+ 108.50	144.93	0.7555	1325	! Q9BKV7 leishmania major. pp9
sp_fungi:Q94317	+ 108.00	152.94	0.6710	534	! Q94317 schizosaccharomyces pom
sp_fungi:Q96WV6	+ 106.50	131.18	1.47	3971	! Q96WV6 schizosaccharomyces pom
sp_bacteriap:Q53278	+ 106.00	149.96	0.9986	526	! Q53278 mycobacterium tuberculosis
sp_bacteriap:Q97YV0	+ 105.50	143.96	1.26	902	! Q97YV0 sulfolobus solfataricus
sp_plant:Q9VND2	+ 104.00	144.01	1.60	704	! Q9VND2 arabidopsis thaliana (mc
sp_plant:Q9WBD0	+ 103.50	147.39	1.99	458	! Q9WBD0 prunus pyrifolia (japanes
sp_bacteria:Q93MG7	+ 103.50	143.27	1.77	701	! Q93MG7 thiobacillus ferrooxidan
sp_invertebrate:Q13131	+ 103.00	141.78	1.99	754	! Q13131 oncorhynchus mykiss (rai
sp_invertebrate:Q13132	+ 102.00	140.22	2.43	754	! Q13132 oncorhynchus mykiss (rai
sp_bacteriap:Q84233	+ 101.00	144.43	2.57	415	! Q84233 chlamydia trachomatis. n
sp_bacteriap:Q97M42	+ 100.50	147.96	2.55	266	! Q97M42 clostridium acetobutylic
sp_bacteriap:Q83182	+ 100.50	139.78	3.13	619	! Q83182 treponema pallidum. alph
sp_fungi:Q74851	+ 99.50	132.99	4.42	1131	! Q74851 schizosaccharomyces pom
sp_invertebrate:Q17585	+ 99.00	117.53	6.94	4845	! Q17585 mus musculus (mouse). u
sp_bacteriap:Q67347	+ 98.00	137.88	4.92	503	! Q67347 caenorhabditis elegans
sp_invertebrate:Q9Y076	+ 98.00	132.55	5.61	764	! Q9Y076 aquifex aeolicus. hypoth
sp_invertebrate:Q9Y075	+ 97.50	139.74	5.09	873	! Q9Y075 leishmania major. prot
sp_bacteria:Q88779	+ 97.50	139.27	5.15	402	! Q88779 yersinia pestis. transp
sp_bacteriap:Q92DR2	+ 97.50	135.31	5.68	605	! Q92DR2 rickettsia prowazekii. r
sp_plant:Q9ZQF1	+ 97.50	134.97	5.73	627	! Q9ZQF1 arabidopsis thaliana (mc
sp_plant:Q9W2S4	+ 97.50	134.13	5.85	684	! Q9W2S4 arabidopsis thaliana (mc
sp_bacteriap:Q98PQ2	+ 97.50	131.60	6.23	888	! Q98PQ2 mycoplasma pulmonis. iso

sp_bacteriap:O25577 + 97.50 129.66 6.54 1085 ! O25577 helicobacter pylori
sp_archaeap:O28239 - 97.00 138.63 5.67 396 ! O28239 archaeoglobus fulgid
sp_bacteriap:O66973 + 97.00 136.18 6.03 510 ! O66973 aquifex aeolicus. l-
sp_archaeap:Q97BF8 + 97.00 135.79 6.09 531 ! Q97BF8 thermoplasma volcani
sp_bacteriap:O32582 + 97.00 134.45 6.29 610 ! O32582 escherichia coli. ta

seq_name: sp_archaea:Q9HHB5

seq_documentation_block:

ID Q9HHB5 PRELIMINARY; PRT; 364 AA.

AC Q9HHB5;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE ALPHA-GALACTOSIDASE.

GN GALA.

OS Pyrococcus furiosus.

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

OX NCBI_TaxID=2261;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verhees C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF195244; AAC28455.1;

DR InterPro; IPR004300; Glyco_hydro_57.

DR Pfam; PF03055; Glyco_hydro_57; 1_

SQ SEQUENCE 364 AA; 41545 MW; 3E8E0AF5BDCCF2A5 CRC64;

alignment_scores:

Quality: 1532.00 Length: 363

Ratio: 4.441 Gaps: 0

Percent Similarity: 95.041 Percent Identity: 79.063

alignment_block:

US-09-886-400-3 x Q9HHB5

Align seg 1/1 to: Q9HHB5 from: 1 to: 364

1 TTGAGAGCGCTGCTTTTCACGGCAACCTCCAGTATCCGAAATCCCAAA 50
1 MCAAGTALALeulValPheHisGlyAsnLeuGlnTrpAlaGluLeuProly 17
51 GAGCGAAATCCCAAGGTCATAGAGAGCGATACATCCAGTATCCGAGA 100
17 sserGluileProlysValIleGluLysAlaTrpPheProThrIleSerG 34
101 CACTGATTAAAGAGAAATTCCTTTGGGCTCAACATAAGGGCTATACC 150
34 luleuileArgargGluileProPheGlyLeuAsnIleThrGlyTrSer 50
151 TTAAAGTTCCTCCCGAAGGATATTATAGACCTCGTTAAAGGGGCGATCGC 200
51 LeuSerPheLeuProlysAspIleAlaLeuileLysGluGlyIleG1 67
201 GAGTACCTGTAGAGATAATTCGAGACAGTACAGCAGCAGCATATACC 250
67 uSerGlyLeuileGluileLeuGlyThrSerTyThrHisAlaileLeup 84
251 CCCTCCTCCGCTTAGCAGAGTAGAAGCAACAGTTCAGAGAGATAGGGA 300
84 roLeuLeuProleuSerArgValGluAlaGlnIleLysArgAspArgGlu 100
301 GTTAAGGAGAGCTCTTCGAGCTTTCTCCAAAGGATTCTCGCTGCAGAGA 350
101 VallyLysGluAsnIleLeuGluValSerProGluGlyPheTrpLeuProG1 117
351 GCTCGCCTATGACCGAATATCCCTGCCATATCTAGAGACACACGGTTATG 400
117 uLeuAlaTrpAspProIleIleProIleAlaileLeuArgAspAsnAsnTrpG 134
401 AGTATCTATTCGCCGACGGGGGCGATGCTTTTCTCAGTCTCATCTCAAC 450

|||||
134 lutyrluPheAlaAspGlyGluAlaMetLeuPheSerAsnHisLeuAsn 150
451 TCSCGCGATAAGCAATTAACCGCTCTATCCACACCTTATAAGGCCCA 500
151 SerAlaileysProlelyProLeuTyrrProHisLeuileLysAlaGl 167
501 AAGGAAAAGCGTTTATGATACATACAGCTATCTCCCTGCTCTCAGGGAGC 550
167 nArgGlyGluGlyLeuValTyrrLeuAsnTyrrLeuLeuGlyLeuArgGlu 184
551 TTAGGAGGCGATAAAGCTCGTTTGAAGCTTAAGCTAAGCTAAAGGCA 600
184 euLysLysAlaileAsnLeuValPheGluGlyLysValThrLeuGluAla 200
601 GTCAAAGACATCGAAGCGCTPACCGCTTGGTGGCGGTGAACACAGCGCTGT 650
201 ValLysGluileGluAlaileProValTrpValSerIleAsnThrAlaVa 217
651 AATGCTCGGATCGGAAGCTTCTCTCTTATGATCTCTAAGAAAGTGGCGA 700
217 lMetLeuGlyAlaGlyArgPheProLeuMetAsnProLysLysValAlaL 234
701 GCTGATAGAGCAAGCAAGCAATCTTCTATACGGCACCGCATATAGAG 750
234 yStrpValLysGluLysAspGluileLeuLeuTyrrGlyThrAspIleGlu 250
751 TTCATTGGCTATAGGACATTCGAGCTACAGNATGAGTGTGAGGGATT 800
251 PheLeuGlyTyrrArgAspIleAlaGlyTyrrLysIleThrIleSerAsnLe 267
801 ATTAGAGGTTATAGACAGCTCAACTCGGAACCTGTGCTCTCCCTCAGAGC 850
267 uLeuGluileleAsnGluLeuGluGlyLeuGlyLeuProArgLysI 284
851 TGAAGCACATGGAAGGAGCTCTACTACGGACTTCGAGTTCGAGTTCGACCA 900
284 lLysHisSerGluLysLysLeuTyrrLeuArgThrSerTrpAlaPro 300
901 GATAAGAGCTTGAGTATATGAGGAGGAGGAGCAAGGAAGCAAGCAAGCTTAA 950
301 AspLysSerLeuArgIleTrpThrGluAspGluGlyAsnAlaArgLeuAs 317
951 TATGCTCTCTACATATAGGAGGAGGAGCTCGCCCTTTAGCGGAGCA 1000
317 nMetLeuThrSerTyrrMetAspGlyGluLeuAlaPheLeuAlaGluAsnS 334
1001 CGGATGCAAGGGGATGGAGCCCTCCCTCAGAGAGGAGGCTGGATGCTTC 1050
334 erAspAlaArgGlyTyrrPgluProLeuProGluArgLeuAspAlaPhe 350
1051 CGGCGGATATATACGATTGGAGGGGTGAATGGGAA 1089
351 LysAlaileTyrrHisTrpArgSerGluAsnGlyLys 363

seq_name: sp_archae:058106

seq_documentation_block:
ID 058106 PRELIMINARY; PRT; 364 AA.
AC 058106;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 41.8 KDA PROTEIN PH0368.
GN PH0368.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,

RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000002; BAA29442.1;
DR InterPro; IP004300; Glyco_hydro_57;
DR Pfam; PF03065; Glyco_hydro_57;
KW Hypothetical protein; Complete proteome
SQ SEQUENCE 364 AA; 41755 MW; 7B4536AB4A975BAD CRC64;

alignment_scores:
Quality: 1494.00 Length: 363
Ratio: 4.381 Gaps: 0
Percent Similarity: 93.939 Percent Identity: 75.758

alignment_block:
US-09-886-400-3 x 058106

Align seg 1/1 to: 058106 from: 1 to: 364

1 TTGAGAGCGCTCGCTCTTTCACGGCAACCTCCAGTATGCCGAAATCCCAA 50
1 MetArgAlaLeuilePheHisGlyAsnLeuGlnTyrrAlaGluileProly 17
51 GAGCGAAATCCCAAGTGCATAGAGAGGATACATCCAGTCATCGAGA 100
17 sHisGluileSerLysValileGluLysSerTyrrPheProThrIleSerG 34
101 CACTGATTAAAGAAAGAAATTCCTTTGGGCTCAACATAACGGGCTATACC 150
34 lueuileLysArgGluileProPheGlyLeuAsnIleThrGlyTyrrSer 50
151 TTAAGTCTCCCGAAGGATATATAGACCTCGTTAAAGGGGCGATCGC 200
51 LeuGlnPheLeuProGlnGluLeuileHisLeuileLysGluGlyileG 67
201 GAGTGACCTCATAGATATATCGGAACGAGCTACACGACGCAATACTCC 250
67 uSerGluileleGluileuGlyThrSerTyrrHisAlaileLeuP 84
251 CCTCTCCCGCTTAGCAGAGTAGAAGCAAGTTCAGAGAGATAGGAA 300
84 roLeuLeuThrLeuSerArgileGluAlaGlnileLysArgAspArgGlu 100
301 GTTAAGGAAGAGCTCTCGAGCTTCTCCAAAGGGATTCGGCTCCAGA 350
101 ileLysGluilePheGluValSerProGlyGlyPheTrpLeuProgl 117
351 GCTCGCTATGACCCGATATCCCTGCCATCTGAAAGGACACGTTATG 400
117 uLeuAlaTyrrAspProIleleProAlaileLeuArgAspAsnGluTyrg 134
401 AGTATCTATTCCGCGGAGCGGCGATGCTTCTCAGCTCATCTCAAC 450
134 lutyrluPheAlaAspGlyGluAlaMetLeuPheSerAsnHisLeuAsn 150
451 TCSCGCGATAAGCAATTAACCGCTCTATCCACACCTTATAAGGCCCA 500
151 SerAlaileysSerileLysProLeuTyrrProTyrrLeuileLysAlaGl 167
501 AAGGAAAAGCGTTTATGATACATACAGCTATCTCCCTGCTCTCAGGGAGC 550
167 nArgGlyGluGlyPheValTyrrLeuAsnTyrrLeuGlyLeuArgGlu 184
551 TTAGGAGGCGATAAAGCTCGTTTGAAGCTTAAGCTAAGCTAAAGGCA 600
184 euLysLysAlaileAsnLeuThrPheGlyGlyLysValThrLeuGluAla 200

```

601 GTCACAGACATCGAACCGCGTACCGCTTTGGTGGCGGTGACACACGCGTGT 650
|||||
201 VallysaspilleGluAlaileProValTrpValSerIleAsnIleAlaI 217
|||||
651 AATGCTGGGATCGGAAGGCTTCTCTATGAATCCATAGAAAGTGGCGA 700
|||||
217 eMeLeuGlyAlaGlyArgPheProLeuMetSerProLysArgValAlaA 234
|||||
701 GCTGATAGAGACAGACAGCAACATCTTCTATACGGCACCGCATATAGAG 750
|||||
234 snTrpIleLysGlyLysAspGluIleLeuLeuTrpGlyThrAspIleGlu 250
|||||
751 TTCATTGCTATAGGACATGCTCAGGCTACAGATGAGTGTTCAGGGATT 800
|||||
251 PheLeuGlyTrpArgSerIleAlaGlyHisLysIleAlaIleSerGlyLe 267
|||||
801 ATTAGAGTTATAGACAGAGCTCAACTCGGAAGTGTGCTTCCCTCAGAGC 850
|||||
267 uMetGluValLeuAsnGluLeuGlyGlyGluLeuCysLeuProArgAspI 284
|||||
851 TGAAGCAGATGGAGGAGGCTCTACTTACGGACTTCGAGTTCGAGTGGCACCA 900
|||||
284 leArgHisAsnGlyArgLeuTrpLeuArgThrSerSerTrpAlaPro 300
|||||
901 GATAAGAGCTTGAGGATGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGCTTAA 950
|||||
301 AspLysSerLeuArgIleTrpLysGluAspGluGlyAsnAlaArgLeuAs 317
|||||
951 TATGCTGTCTACATATAGAGGGCGAACTCGCGCTTTTAGCCGAGAACAA 1000
|||||
317 nMetLeuThrTrpCysMetAspGlyGluPheAlaPheLeuAlaGluAsnS 334
|||||
1001 CGGATGCAAGGGATGGAGCGCCCTCCCTGAGAGAGGCTGGATGCTTC 1050
|||||
334 eRAspAlaArgGlyTrpGluProLeuProGluArgArgLeuAspAlaPhe 350
|||||
1051 CGGCGATATATAAGTGTGAGGGGTCAAATGGGGAA 1089
|||||
351 LysAlaIleTrpLysTrpArgAsnGluAsnGlyLys 363

```

seq_name: sp_bacteriap.74630

```

seq_documentation_block:
ID P74630 PRELIMINARY; PRT; 529 AA.
AC P74630;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HIPOTHETICAL 62.1 KDA PROTEIN.
GN SL0735.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kareko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90916; BAA18743.1;
DR InterPro; IPR003803; DUF200.
DR Pfam; PF02651; DUF200; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 529 AA; 62075 MW; 4C1A45048A784E30 CRC64;

```

alignment_scores:

```

Quality: 162.50 Length: 471
Ratio: 0.860 Gaps: 19
Percent Similarity: 40.127 Percent Identity: 21.444

alignment_block:
US-09-886-400-3 x P74630 ..

Align seg 1/1 to: P74630 from: 1 to: 529

7 CGCGTCGCTCTTCACGGCAACCTCCAGTATCGGAAATCCCAAGAGCGGA 56
|||||
7 AlaLeuValLeuHisAlaHisLeuProPheValArgHisProGluSerAs 23
57 A.....ATCCCAAAGTCTATAGAGAGGCGATACA 85
|||||
23 pTyrValLeuGluGluTrpLeuTrpGluAlaIleThrGluThrTrpI 40
|||||
86 TCCAGATCATC.....GAGACACTGATTAAAGAGAAATTCCTTTT 126
|||||
40 leProLeuIleHisValPheGluGlyLeuLysArgAspGlyValAspPhe 56
|||||
127 GGGCTCAACATAACG..... 141
|||||
57 LysIleThrMetSerMetThrProProLeuValSerMetLeuArgAspPr 73
|||||
141 ..... 141
|||||
73 oLeuLeuGlnGlnArgTrpGluAlaHisLeuSerLeuLeuGlnGluLeuL 90
|||||
142 .....GGCTATACC 150
|||||
90 euAlaLysGluIleValArgAsnGluHisAsnGlyHisLeuGlnTrpLeu 106
|||||
151 TTAAGTCTCTCCCGAAG..... 168
|||||
107 AlaAspPheTrpAlaLysGluPheAlaAlaIleArgGluThrTrpGluAr 123
|||||
169 .....GATATTATAGACCTCGTTAAAGGGGCGCATCGCAGGTGACC 208
|||||
123 gTyRAspGlyAspLeuValThrAlaPheLysGlnPheGlnAspSerAsnA 140
|||||
209 TGATAGAGATATCGGACGAGCTACAGCGACCAATACTCCCGCTCCTC 258
|||||
140 snLeuGluIleIleThrCysGlyAlaThrHisGlyTrpPheProLeuMet 156
|||||
259 CGGCTT.....AGCAGATAGAACACAAAGTTCAGAGAGATAGGGAAGT 302
|||||
157 LysMetTrpProGlnAlaValTrpAlaGluIleLysValAlaLysGluHi 173
|||||
303 TAAGAGAGAGCTCTTCGAGCTTCTCCAAAGGGATTCGGCTGCCAGAGC 352
|||||
173 sTyRgLuGluAsnPheGlyArgSerProLysGlyIleTrpLeuProGluC 190
|||||
353 TCGCTATGACCCGATATACCTCCCATCTGAAGGACAAACGGTATTAG 402
|||||
190 ysAlaTrpTrpGluGlyValGluArgMetLeuAlaAspAlaGlyLeuArg 206
|||||
403 TATCTATTTCGCGCGGAGGAGGATGCTTTTCTCAGCT..... 441
|||||
207 TyrPheLeuValAspGlyHisGlyIleLeuTrpAlaArgProArgProLy 223
|||||
441 ..... 441
|||||
223 sPheGlyThrTrpAlaProIlePheThrGluThrGlyValAlaAlaPheG 240
|||||
442 .....CATCTCAACTCG.....GGGATAAAG 462
|||||
240 lyArgAspHisGluSerSerGlnGlnValTrpSerSerGlnValGlyTy 256
|||||
463 CCATTAACCGCTCTATCCACACCTTATAAAG...GCCCAAGGGGAAA 509
|||||
257 ProGlyAspProValTrpArgGluPheTrpLysAspLeuGlyTrpGluAl 273

```



```

510 GCCTTTAGTACATCAGCTATCTCTGGTCTCAGGAGGCTTAGAAGG 559
273 aGlutyrgluTyrlleLysProTyrileMetProAsnGlyGlnArgLysA 290
560 CGATAAAGCTCTCTTTT.....GAAGGTAAGGTAACG 591
290 snilegilyleLysTyrHisLysileThrSerArgAspGlyGlyLeuSer 306
592 CTAAGGCGAGTCAAGACATCGAACCGCTACCGCTTGGTGGCGGTGAA 641
307 gluLysAlaItrpTyrAsp.....ProTyrTrpAlaLysGluLy 319
642 CACGGCTGTA..... 651
319 sAlaAlaGluHisAlaSerAsnPhenMetTyrAsnArgGlnGlnValG 336
652 .....ATGCTCGGATC...GGAAGCTCTCTTATGAATCCTAAGAAA 693
336 lyHisLeuSerGlyileMetGlyArgProLeuValValSerProTyr 352
694 GTGGCGAGC.....TGGATAGAG..... 711
353 AspAlaGluLeuPheGlyHisTrpTrpTyrGluGlyProTyrPheIleAs 369
712 .....GACAGGACAAACATT..... 726
369 pTyrLeuPheArgLysSerTrpPheAspGlnAspThrPheGluMetThr 386
727 .....CTTCTATACGCCACCGATATAGAGTTTCATT..... 756
386 isLeuAlaAspTyrLeuArgGlyAsnProHisGlnGlnValCysArgPro 402
757 .....GGCTATAGGACATTCAGGCTAC..... 780
403 SerGlnSerSerTrpGlyTyrLysGlyPheHisGluTyrTrpLeuAsnAs 419
781 .....AGAATGA 787
419 pThrAsnAlaTrpIleTyrProHisLeuHisLysAlaAlaGluArgMetI 436
788 GTGTTGAGGATATTAGAGTTATAGACGAGCTCAACTCGAAGCTGTGC 837
436 leGluLeuSerHisArgGluAlaValAspGluLeuGluGluLys..... 450
838 CTTCCCTCAGAGCTGAAGCACAGTGGAGGAGGCTCTACTTA...CGGAC 884
451 .....AlaLeuAsnGlnAlaAlaArgGluLeuLeuAlaGlnse 464
885 TTCGAGTGGGCA 897
464 rSerAspTrpAla 468
seq_name: sp_archaeap:Q97BM4
seq_documentation_block:
ID Q97BM4 PRELIMINARY; PRT; 378 AA.
AC Q97BM4;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE ALPRA-AMYLASE.
GN TVG0421416.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
OC Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GSS1 / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunoshiba T., Yamamoto Y., Atamaki H., Makino K., Suzuki M.;

```

```

RT "Archaeal adaptation to higher temperatures revealed by genomic
RL sequence of Thermoplasma volcanium."
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
DR EMBL; AP00992; BAB59573.1; -.
KW Complete proteome.
SQ SEQUENCE 378 AA; 44957 MW; 49FCF63BEB6B3A7 CRC64;

alignment_scores:
Quality: 149.50 Length: 313
Ratio: 1.017 Gaps: 11
Percent Similarity: 46.965 Percent Identity: 21.086

alignment_block:
US-09-886-400-3 x Q97BM4 ..
Align seg 1/1 to: Q97BM4 from: 1 to: 378

64 AAGGTCATAGAGAGCATACATCCAGTCATCGAGACACTGATTAAAGA 113
42 ArgValAlaGluArgSerTyrIleProAlaThrArgAsnLeuMetGluTy 58
114 AGAAATTCCTTTTGGCTCAACATAACGGC..... 144
58 rGlyIleLysSerSerPheSerIleThrGlyThrAlaValGlnAlaL 75
145 .....TATACCTTAAAGTTCCTCCGAGGATATTATAGACCTCGTTAA 189
75 euMetTyrAsnThrLys.....ValIleAspAlaIleAsp 86
190 GGGGCGATCGCGAGTGACGTATAGATAGATAATCGAAGCAGCTACACGA 239
87 AspTyrValLysSerGlyLeuCysGluMetLeuSerGluThrTyrTrpHi 103
240 CGCAATACCTCCCTCCCTCCGCTTAGCAGATAGAGACACAACTCAGA 289
103 sSerLeuAlaSerIleTrpAsnTyrAspGluPheLysArgGlnValAspM 120
290 GAGATAGGGAAGTTAAGGAGAGCTCTTCGAGCTTCTCCAAAGGGATTC 339
120 eHisArgAspLeuMetLysArgIlePheAsnValValProLysValPhe 136
340 TGCGTCCAGAGCTGCGCTATGACCCGATAATCCCTGCCATCTACTGAAG 389
137 ArgAsnThrGluLeuIleTyrAspAspArgIleAlaGluMetValLysAr 153
390 CRACGGTTATAGTATCTATTCGCCGACGGGAGCGGATGCTTTCTCAG 439
153 gMetGlyPheThrSerIleIleThrGluGly..... 163
440 CTCATCTCAACTCGGCGATAAAGCAATTAAACCGCTCTATCCACACCTT 489
164 .....ThrAspSerIleValLysAspHisSerProAsnTyr..... 175
490 ATAAAGGCCCAAGGAAAGCGCTTTAGGTACATCAGC.....TA 530
176 .....ArgTyrAlaSerProSerGlyLe 183
531 TCPCCTTGCTCAGGAGCTTAGGAAGCGCATAAAGCTCGTTTGAAG 580
183 uAsnLeuTyrLeuArgAsnTyrValMetSerAspAsnIleSerPheArgP 200
581 GTAAGTAACGCTAAAGGCGAGTCGAAGACATCGAAGCGCTACCGGTTGG 630
200 heSerAsnThr.....LysTrpLysAspTyrPro..... 209
631 GTGGCGGTGAACACAGGCTGTATGTCGGCATCGAAGGCTTCCTCTTAT 680
210 .....Le 210
681 GAATCCTTAAGAAAGTGGCGAGCTGGATAGAGGACAAAGGACACATCTTC 730
210 uThrAlaAspLysTyrAlaLysTrpIleAsnGluSerGluGlyAspMetV 227

```

```

RT "Archaeal adaptation to higher temperatures revealed by genomic
RL sequence of Thermoplasma volcanium."
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
DR EMBL; AP00992; BAB59573.1; -.
KW Complete proteome.
SQ SEQUENCE 378 AA; 44957 MW; 49FCF63BEB6B3A7 CRC64;

```

```

731 TATAGCGACCGATATAGAGTTTCATGGCTATAGGACATTCGAGCTAC 780
  ::      |||::: |||::: |||::: |||::: |||::: |||:::
227 alasn.....LeuPheMetAspTyrGluThrPheGlyGluHis 239
  ::      |||::: |||::: |||::: |||::: |||::: |||:::
781 AGAATGAGTGTGAGGATATTAGAGTTATAGACGAGCTC..... 822
  ::      |||::: |||::: |||::: |||::: |||::: |||:::
240 GlnThrGlnGluThrGlyIlePheAspPheMetLysTyrLeuProValTy 256
  ::      |||::: |||::: |||::: |||::: |||::: |||:::
823 .....AACCTGGAACTGTGCCTTCCTCCCTCAGACTGA 853
  ::      |||::: |||::: |||::: |||::: |||::: |||:::
256 rPheArgAspTyrGlyIleGluThrIleThrIleSerGluAlaGluLysA 273
  ::      |||::: |||::: |||::: |||::: |||::: |||:::
854 AGCACAGTGGAGGAGCTCTACTTA.....CGGACTTCGAGTTGGGCA 897
  ::      |||::: |||::: |||::: |||::: |||::: |||:::
273 rGHisArgValLysAspValLeuSerIleProGluThrIleSerIlePala 289
  ::      |||::: |||::: |||::: |||::: |||::: |||:::
898 CCAGAT...AAGACCTTGAGGATGATGAGAGGAGGAGAA 933
  ::      |||::: |||::: |||::: |||::: |||::: |||:::
290 AspThrArgArgAspLeuSerAlaTrpLeuGluAsnGlu 302
  ::      |||::: |||::: |||::: |||::: |||::: |||:::
seq_name: sp_bacteriap:Q97GF3

seq_documentation_block:
ID   Q97GF3          PRELIMINARY;      PRT;       527 AA.
AC   Q97GF3;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNCHARACTERIZED CONSERVED PROTEIN, RELATED TO
DE   ALPHA-AMYLASE/ALPHA-MANNOSIDASE.
GN   CAC2414.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
CX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum."
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL: AE007742; AAK80369.1;
DR   InterPro: IPR003803; DUF200.
DR   Pfam: PF02651; DUF200; 1.
KW   complete proteome.
SQ   SEQUENCE 527 AA; 61485 MW; 03706E16907929C6 CRC64;

```

```

alignment_scores:
  Quality: 147.00      Length: 486
  Ratio: 0.766        Gaps: 20
  Percent Similarity: 39.506      Percent Identity: 18.930

alignment_block:
US-09-886-400-3 x Q97GF3
..
Align seg 1/1 to: Q97GF3 from: 1 to: 527

13 GTCITTCAGCGCACCTCCAGTATGCGGAATCCCAAGAGCGAA..... 57
9 ValLeuHisSerHisMetProPheValArgHisProGluThrArgAspSe 25
58 .....ATCCCAAAGGTCATAGAGAGCGCATACATCCAG 91
25 rleuGluGluArgTrpLeuPheGluAlaMetSerGluCysTyrIleProL 42
92 TCATCGAG.....ACATGTATTAAAGAAATTCCTTTTGGG... 129

```

```

:::|||||
42 euileGluValTy rAspAsnLeuLeuLysAspAsnIleLysPheArgMet 58
129 .....|||::: |||::: |||::: |||::: |||::: |||:::
59 ThrMetSerIleThrProLeuMetSerMetLeuGlnAspGluTyrLe 75
130 .....CTCAACATACGGGCTATACCTTAAAGTTCCTCCCGA 166
75 uAsnSerArgTyrLeuAsnTyrLeuLysTyrThrIleGluLeuSerGluL 92
167 AGGATATTATA.....|||::: |||::: |||::: |||::: |||::: |||:::
92 ySGluIleLeuArgThrLysAsnAsnArgGluGluAsnLysValAlaLeu 108
177 .....|||::: |||::: |||::: |||::: |||::: |||:::
109 PheTy rAsnLysArgAlaGluAsnThrLeuLysIleTyrGluLysTyrAs 125
178 ....GACCTCGTTAAAGGGGCATCGCGAGTGAC.....CTGATAG 214
125 pAsnAsnLeuIleAsnAlaPheArgLysTyrAspArgLeuGlyCysValG 142
215 AGATAATCGGAACGAGCTACACGCACCAATACCTCCCTCTCCCGCCT 264
142 luileIleThrCysAlaAlaThrHisAlaLeuLeuProLeuIleLeuIle 158
265 AGCAGA.....GTAGAAGCAGAGTTCAGAGATAGGAGACTTAAGGA 308
159 AsnArgGlnAlaValLysAlaGlnIleAlaThrGlyValGlnSerTyrIl 175
309 AGAGCTCTTCGAGCTTTCTCCAAAGGATTCTGGCTGCCAGAGCTCGCCT 358
175 eAsnThrMetGlyHisGluProAsnGlyIleTrpLeuProGluCysAlat 192
359 ATGACCCGATATCCCTGCCATACCTAGGACACACGGTTATGAGTATCTA 408
192 YrThrTy rGlyIleAspAsnIleLeuSerGluPheGlyIleLysTyrPhe 208
409 TTCGCCAGCGGAGGAGGATGCTTTCTCAGCTCATCTCAACTCGCGCAT 458
209 IleSerGluGlyLysAlaIleAspTyrAlaSerPro.....220
459 AAAGCCAAATTAACCGCTCTATCCACACCTTATAAGGCCCAAGGAAA 508
221 .....LysProMetTyr.....224
509 AGCGCTTTAGGTACATCAGCTATCTCCTTGGTCTCAGGAGCTTAGGAAG 558
224 .....224
559 GCGATAAAGCTCGTTTTTGAGGTAAAGTAACGCTAAAGCAGCTCAAGA 608
224 .....224
609 CATCGAAGCCGTACCGCTTTGGTGGCGGTGAACACGGCTGTATGCTCG 658
225 .....GlyThrAsnThr.....228
659 GCATCGGAAGGCTTCCTCTTATGATTCCTTAAGAAAGTGGCGAGCTGATA 708
229 .....ProIleAlaAlaProSerGlyValCysAlaPheGly 240
709 GAGGACAGGAGCAACATTCCTTATACGGCAGCGATATAGATTTCATTGG 758
241 ArgaspMetaspSer.....SerTy rGlnValTrpSeraspPheMetGl 255
759 C.....TATAGGACATTGCAGGCT 778
255 Ty rProGlyaspPheAsnTy rArgGluPheTy rArgAspile...GlyP 271
779 ACAGATGAGTGTTCAGGAGTATTATAGAGGTATTAGACGAGCTCAACTCG 828
:::|||||

```



```

993 CGAAGACGCGATCAAGGGGATGGAGCCCTC.....CCTGAGGGA 1036
696 YAspAnilleSerProLeuValtyrLeuProPheAsnThrThrProAsnS 713
1037 GGCTGGATCCCTCCGGGGATATATAC..... 1065
713 erThrAspLeuIleAspThrLeuTyrAsnTyrLeuTyrValalaGluGly 729
1066 ...GATTGGAGGGTCAAAATGGGAACCT 1092
730 SerAspTyrThrTyrPheGlnThrGlyProPro 739
seq_name: sp_archaeap:050094

seq_documentation_block:
ID O50094 PRELIMINARY; PRT; 560 AA.
AC O50094;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE HYPOTHETICAL 65.8 KDA PROTEIN PH1386.
GN PH1386
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 3:53-76(1998).
DR EMBL; AP000006; BAA30492.1; -.
DR InterPro; IPR003803; DUF200.
DR Pfam; PF02651; DUF200; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 560 AA; 65769 MW; 3F4CE5267D167F8 CRC64;

```

```

alignment_scores:
  Quality: 130.50      Length: 234
  Ratio: 1.186        Gaps: 12
  Percent Similarity: 47.009 Percent Identity: 25.641

alignment_block:
US-09-886-400-3 x O50094 ..
Align seg 1/1 to: O50094 from: 1 to: 560

40 GAAATCCCAAGACGGAATCCCAAGGTCATAGAGAGGATACATCC 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71 GlutrylleLysargPheGluLysTyrMetGluArgLysLeuLys 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 AGTCATCGAGACACTG.....ATTAAAGAGAAA 118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
87 rMetGluGluAspLeuGluArgPheLysAspGluLysLeuArgGluAlaI 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
119 TTCTTTTGGGCTCAACATAACGGGTATACCTAAAGTTCCTCCCGAAG 168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104 leasnPhe.....MetIleGlyTyrPhe.....Lys 112
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 GATATTATAGACCTC.....GTTAAGGGGCGATCGGAGTGAACCT 209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
113 AspValTyrSerTyrTrpLysSerIleAspGlyAsnIleLeuGlyLysPh 129
210 G.....ATAGAGATAATCGGAACGAGCTACA 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

129 eArgGluLeuGlnAspGluGlyTyrValGluValIleThrSerAlaAlaIat 146
236 CCGACGCAATACTCCCTCCCTCCTCCGCTTATGACAGA...GTAGAAGCACAA 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
146 hrHisGlyTyrLeuProLeuLeuGlyArgAspGluAlaIleGluAlaGln 162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 GTTCAGAGAGATAGGAAAGTTAAGGAAGAGAGCTTCGAGAGCTTTCCTCAA 332
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
163 LeuLeuAsnGlyIleLysValTyrGluLysTyrPheGlyArgLysProAr 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
333 GGGATTCTGGCTCGCAGAGCTCGCCTATGACCCGATA..... 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
179 gGlyIleTrpLeuProGluCysAlaTyrArgProAspGlyLeuTrpLys 196
370 .....ATCCCTGCCCATCTG 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
196 erProSerThrGlyGluValLysTyrPheGlyIleGluHisPheLeu 212
385 AAGGACACGGTTAGTATATCTATTCGCCGAGCGGGAGCGATGCTTTT 434
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
213 LysLysPheGlyIleGluTyrPheValGlu..... 223
435 CTCAGCTCATCTCACTCGGCGATTAAGCCAAATTAACCGCTCTATCCAC 484
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
224 ....SerHisLeuIleAspLys...GlyProValSerLeuArgTyrGlyA 238
485 ACCTTATAAAGGCCCAAGGAAAGCGCTTTAGGTACATCAGCTATCTC 534
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
238 snIleLeuProAla...LysThrLysArgSerThrLeuArgProTyrPhe 253
535 CTGGTCTCAGGAGCTTAGGAGCGGATAGCTCGTTTGAAGGTAA 584
254 .....LeuLysasnGlyIleAlaValPheAlaAa-gasnAr 265
585 GGTAAAGCTAAAGGCGAGTCAAAAGACATCGAAGCCGCTGTTGGTGG 634
265 gGluThr.....GlyIleGlnValTrpSerA 274
635 CC 636
274 la 274

seq_name: sp_bacteriap:09KD04
seq_documentation_block:
ID O9KD04 PRELIMINARY; PRT; 923 AA.
AC O9KD04;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BH1415 PROTEIN.
DE BH1415
GN BH1415
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP01512; BAB05134.1; -.
DR InterPro; IPR003803; DUF200.
DR Pfam; PF02651; DUF200; 1.
DR Pfam; PF00534; Glycos_transf_1; 1.
KW Complete proteome.
SQ SEQUENCE 923 AA; 106638 MW; 8C7B7D64F3603553 CRC64;

```

```

250      ...   |||   |||   |||
257      ...TyrProGlyAspValAspTyrAr
597  GGCA...GTCAAAGACATCGAAGCGGTACCGGTTGG...    630
      :      ||| ||| |||
257  gGluPheTyrArgAspIleAlaTyrAspArgGluTyrAspTyrIleLysP 274
631      ...      GTGGCGGTGACACGGCTGAATGCTCGCGC 660
274  roHisValHisLysAspGlyIleArgIleAspThrGly...LeuLys 288
661  ATCGGAGGCGTCTCTTATGAATCCTAAGAAA...GTGGCGAG 701
289  TyrHisArgIleThrGlyHisThrGluGluLysAspLeuTyrValArgG 305
702  CTGGATAGAGGACAG...
305  utrPalagIuLysArgValGlnGluHisAlaAsnHisPheIleGlyAla 322
718      ...      GACAAACATT 726
322  leHisHisGluIleAspGlnHisGlyGlyGlnAsnPheProProTyrVal 338
727  CTCCTATACGGCACCATATAGATTCATTGGCTATAGGACATTCGAG 776
339  MetValThrProPheAspAlaGluLeuPheGlyHisTrpTrpPheGluG 355
777  CTACAGAAATCAGTGTTCAGGGATTATTAGAGTT...ATAG 814
355  yProGluTyr...IleGluAlaLeuTyrGluGlnGlyAlaAspArgVal 371
815  ACGAGCTCACTCGGACGTGCTTCCTCTAGAGCTGAACACAGTCGGA 864
371  erPheIleThrProGluLeuTyrLeuGlnArgHisTyrGln...AspPhe 386
865  AGGAGGCTCTACTTCGGACTTCGAGTTGGGCACCATAGAGCTTGAG 914
387  GlnThrAlaHisValSerPheSerThrTrpGlyArgAspGlyTyrGlyHi 403
915  GATATGGAGAGAGCAGGACGAGGACCAAGACTTAATATGCTGTCTCTACA 964
403  sValTrpLeuAsnAspHis...AsnAlaTrpMetTyrArgHisTyrHisA 419
965  ATATGAGGGCGGAACTCGCCCTTTA...
419  rgMetGluLysAspLeuAlaLysIleValAlaMetTyrProGlnProThr 435
991      ...GCCGAGAACAGCGATGCAAGGGATGGAGCGCCCT 1025
436  ValLeuGluLysGlnAlaIleGlnGlnMetValArgGluTyr..... 449
1026  CCTGTAGAGAGCGCTGTGATGCCCTCCGGCGATATATACGATTGG 1071
450      ...MetLeuAlaValSerSerAspTrp 457

seq_name: sp_fungi:Q9C105

seq_documentation_block:
ID   Q9C105      PRELIMINARY;      PRS;  1236 AA.
AC   Q9C105;
DT   01-JUN-2001 (TREMBLrel. 17, Created)
DT   01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT   01-DSC-2001 (TREMBLrel. 19, Last annotation update)
DE   PUTATIVE GLUCOMYLASE I (ALPHA-1,4-GLUCAN GLUCOSIDASE), EXT
DE   STARCH-DEGRADING ENZYME, BY SIMILARITY TO S. CEREVISIAE STA
DE   CHITINASE FAMILY SIGNATURE.
GN   SPAPB17.04C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomycetes.
ON   NCBI_TaxID=4896;
RX   [1]

```



```

94 ATCGAGACACTGATTAAAGAGAAATTCCTTTTGGGCTCAACATAACGGG 143
   |||||
236 ValGluThrValLeuLysHisGlnMet...TrpLeuLeuAsnHisThr... 250
   |||||
144 CTATACCTTAAAGTTCTCCGGAAGGATATATAGACCTCGTTAAAGGG 193
   |||||
251 .....PheGluHisGluLysIleAsnLeuLeuLeuGly. 262
   |||||
194 GCATCGCGAGTGCCTGATAGAGATATCGGAACGAGCTACACGACGCA 243
   |||||
263 .....AsnGlyAsnValGluValThrValValProTyrThrHisPro 276
   |||||
244 ATACTCCCTCTCTCCG.....CTTAGCAGAGTAGAGCACA 281
   |||||
277 IleGlyProIleLeuAsnAspPheGlyTyrTrpTyrGluAspPheAspAla 293
   |||||
282 AGTCAGAGAGTAGGAGCTTAAGGAAGAGCTCTTC.....GAGC 322
   |||||
293 nValLysLysAlaAsnGluLeuTyrLysGluTyrLeuGlyAlaGlyLysV 310
   |||||
323 TTCTCCAAAGGATTCGTGCTGCCAGAGCTCGCCTATGACCCGATAATC 372
   |||||
310 alThrProLysGlyTyrTrpAlaAlaGluSerAlaLeuAsnAspLysThr 326
   |||||
373 CCTGCCATCTGAAGACACGCTTATGACTATCTATTCGCGACGGG... 420
   |||||
327 LeuGluLeuLeuAlaGlnGlyTyrLysTrpValMetThrAspGlnLe 343
   |||||
421 .....GAGCGCATGCTTTCTCAGCTCATCTCAACTCGCGGATAAAGC 463
   |||||
343 uValLeuGluLysLeuGlyValProLysThrIleGluSerTyrTyrLysP 360
   |||||
464 CA..... 465
   ||
360 rotTrpValAlaGlnPheGlyAspLysLysIleTyrLeuPheProArgAsn 376
   |||||
465 ..... 465
377 HisAspLeuSerAspArgValGlyPheArgTyrAlaGlyMetAsnGlnTyr 393
   |||||
466 .....ATTAACCGCTCTATCCACACCTTATAAGGCCCAAGGAA 508
   |||||
393 rAspAlaValLysAsnPheValGluGluLeuLeuLysIleGlnLysGlnA 410
   |||||
509 AGCGCTTAGGTACATCAGCTATCTCTCTCGTCTCAGGGAGCTTAGGAAG 558
   |||||
410 snTyrAspGlySerLeuValTyrValIleThrLeuAspGlyGluAsnPro 426
   |||||
559 GCGATAAGCTCGTTTGAAGTAAGTAACGCTAAAGCAGTC...AA 605
   |||||
427 TrpGluHisTyrProPheAspGlyLysLeuPheLeuGluLeuTyrAr 443
   |||||
606 AGACATCGAACCCCTACCCGTTTGGGTGGCGCTGAACACGGCTGTA.... 651
   |||||
443 gGlnLeuGluGluLeuGlnLysLysGlyLeuIleArgThrValThrProS 460
   |||||
652 .....ATGCTCGGC.....ATCGGAAGGCTTCCTCTTATG 681
   |||||
460 erGluTyrIleGluMetPheGlyAspLysAlaAsnLysLeuThrProLys 476
   |||||
682 AATCCTAAGAAGTGGGAGCTGTGATAGAGCAAGGACAACTTTCT 731
   |||||
477 MetMetLysArgLeuAspPheThrThrGluAspAsnValAsnAlaLeuLe 493
   |||||
732 ATACGGCACCATATAGAGTTTCATTGGCTATAGGCACATTGCGAGCTACA 781
   |||||
493 uLysAlaLysThrLeu.....GlyGluL 501
   |||||
782 GAATAGGTGTTGAGGATATATAGAGTTATAGACGAGCTCAACTCGGAA 831
   |||||
501 eutYrAspMetValGlyValThrGlu..... 509

```

```

832 CTGTGCTTCCCTCAGAGCTGAAGCACAGTGAAGGAGGAGCTCTACTTAG 881
   |||||
510 .....GluMetGlnTrpPr 514
   |||||
882 GACTTCGAGTTGGCCACAGATAAGAGCTTGAGGATATGAGAGAGGAGCG 931
   |||||
514 oGluSerSerTrpIle...AspGlyThrLeuSerThrTrp.....I 527
   |||||
932 AAGGGAACGCAAGACTTAATATGCTGCTCTCAATATGAGGGCGCAACTC 981
   |||||
527 leGlyGluProGlnGluAsnIleAlaTyrTyrTrpLeuTyrLeuAlaArg 543
   |||||
982 GCCCTTTTACCGGACACAGAT...GCAAGGGGATGGAGGCCCTCC 1028
   |||||
544 LysAlaLeuPheGluAsnLysAspAsnValLysAspTrpAsnLysAlaTy 560
   |||||
1029 TGAGAGAGGCTGATGCTTCGCGGCGATATATAACGATTGG 1071
   |||||
560 rGluTyr.....LeuPheArgAlaGluGlySerAspTrp 571
   |||||
seq_name: sp_archaeap:Q9HL91
seq_documentation_block:
ID Q9HL91 PRELIMINARY; PRT; 357 AA.
AC Q9HL91;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ALPHA-AMYLASE RELATED PROTEIN.
GN TA0339.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Wewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum."
RL Nature 407:508-513(2000).
DR EMBL; AL445064; CAC11483.1; -.
KW Complete proteome.
SQ SEQUENCE 357 AA; 41845 MW; B8C1A97F4EB237F4 CRC64;

```

alignment_scores:

```

Quality: 122.50 Length: 307
Ratio: 0.869 Gaps: 8
Percent Similarity: 45.928 Percent Identity: 16.612

```

alignment_block:

```

US-09-886-400-3 x Q9HL91

```

```

Align seg 1/1 to: Q9HL91 from: 1 to: 357

```

```

64 AAGTCATAGAGAGGATATATAGAGCTATCGAGACACTGATTAAAGA 113
   |||||
20 ArgIleAlaAlaAsnTyrIleProAlaThrGlnSerLeuMetAspTy 36
   |||||
114 AGAATATCCITTTGGGCTCAACATACGCGCTATACCTTA..... 153
   |||||
36 rGlyIleArgSerSerPheSerLeuSerGlyThrPheMetGluGlnAla 53
   |||||
154 ..AAGTTCCTCCGGAAGGATATATAGAGCTCGTTAAAGGGGCGCATCGG 201
   |||||
53 erLysTyrCysProLys...ValIleAspValIleAspTyrValArg 68
   |||||
202 AGTACCTGATAGAGATAATCGGAACGAGCTACACGACCACTACTCC 251
   |||||
69 SerGlyGlnCysGluLeuLeuSerGluThrTyrHisSerLeuAla 85

```



```

252 CTTCTCCCGCTTACGAGAGTAGAAGCACCAAGTTCAGAGAGATAGGGAAG 301
   ::::::::::::::::::::|
85 rIleTPrAsnAspGluGluPheValArgGlnValArgMetGlnGluSerA 102
   ::::::::::::::::::::|
302 TTAAGGAAGAGCTCTTCAGAGCTTCTCCAAAGGGATTCTGGCTGCCAGAG 351
   ::::::::::::::::::::|
102 latLeuLysThrPheAsnTyrGluProValSerPheArgAsnThrGlu 118
   ::::::::::::::::::::|
352 CTGCGCTATGACCGCATATACCTGCCATCATGAAGGACCAACGGTTATGA 401
   ||| |||::: ||| :::: ||| |||::: |||
119 LeuIleTyrAsnAspHisIleAlaGluValAlaLysGlyMetGlyPheAr 135
   ::::::::::::::::::::|
402 GTATCTATTCGGCGAGGGAGCGATGCTTTTCACGCTCATCTCAACT 451
   ::::::::::::::::::::|
135 gAsnIleLeuAlaGluGlyThrAspAspIleAlaSerArgTyr..... 149
   ::::::::::::::::::::|
452 CGCGCATAAAGCAATTAACCGCTCTATCCACACCTTATAAGGCCCA 501
   ::::::::::::::::::::|
149 ..... 149
502 AGGGAAGCGGTTTAGGTAC.....ATCAGCTATCTCT 536
   ::::::::::::::::::::|
150 ...AspValAsnTyrArgTyrAlaAlaProSerGlyIleAsnLeuTyrLe 165
   ::::::::::::::::::::|
537 TGGTCTCAGGAGCTTAGGAAGCGCATAAAGCTCGTTTGAAGCTAAGG 586
   ||| |||::: ||| |||::: ||| |||::: |||
165 uArgAsnTyrProLeuSerAspAspIleSerPheArgPheSerAsnArgA 182
   ::::::::::::::::::::|
587 TAACGCTAAAGCAGCTAAAGACATCGAAGCGGTACCCGTTTGGGTGCC 636
   ::::::::::::::::::::|
182 latrPAlaAspTyrProLeuThrAlaAspLysPheAlaArgTyrPheSer 198
   ::::::::::::::::::::|
637 GTGACAGCGCTGATATCGGCATCGAAGCGCTTCCTCTATCAATCC 686
   ::::::::::::::::::::|
199 AlaSerSerGlyIleValAsnLeu..... 207
   ::::::::::::::::::::|
687 TAAGAAAGTGGCAGCTGGATAGAGGACAAAGCAACATTTCTTATACG 736
   ::::::::::::::::::::|
207 ..... 207
737 GCACCGATATAGATTTCATGGCTATAGGACATTCGAGGCTACAGATG 786
   ::::::::::::::::::::|
208 .....PheMetAspTyrGluThrPheGlyGluHisGlnArg 219
   ::::::::::::::::::::|
787 AGTGTGAGGATTATAGAGTTATA..... 813
   |||::: |||::: |||::: |||::: |||
220 ProGluThrGlyIlePheGluPheLeuArgTyrLeuProMetTyrPheGl 236
   ::::::::::::::::::::|
814 ....GACGAGCTCACTCGAAGCTGTGCTTCCTCCCTCAGAGCTGAAGCACA 859
   ::::::::::::::::::::|
236 uGluAsnAspValHisThrIleLeuValArgGluAlaGluAlaArgHisA 253
   ::::::::::::::::::::|
860 GTGGAAGGGAGCTC.....TACTTACGAGCTTCGAGTTGGSCA...CCA 900
   |||::: |||::: |||::: |||::: |||
253 rgThrArgAspPheIleSerValSerLysThrThrSerTyrAlaAspLys 269
   ::::::::::::::::::::|
901 GATAAGAGCTTGAGGTATCG 921
   ::::::::::::::::::::|
270 AsnArgAspLeuSerAlaTyr 276
   ::::::::::::::::::::|
seq_name: sp_archaeap:030246
seq_documentation_block:
ID O30246 PRELIMINARY; PRT: 324 AA.
AC O30246;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DE 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE METHANOL DEHYDROGENASE REGULATORY PROTEIN (MOXR).
GN AF2425.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;

```

```

OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1];
RP SEQUENCE FROM N.A. 4304 / ATCC 49558;
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Fleischmann D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Raine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AE001108; AAB91247.1; -.
DR TIGR; AF2425; -.
DR InterPro; IPR000523; Mg_chelatase_chII.
DR Pfam; PF01078; Mg_chelatase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 324 AA; 36609 MW; B0F64965EFC82F1F CRC64;

```

alignment_scores:

```

Quality: 120.50 Length: 327
Ratio: 0.748 Gaps: 14
Percent Similarity: 49.235 Percent Identity: 21.713

```

alignment_block:

```

US-09-886-400-3 x O30246

```

```

Align seg 1/1 to: O30246 from: 1 to: 324

```

```

1 TTGAGAGCGCTGCTCTTTTCACGGCAACCTCCAGTATGCCGAATCCCA... 48
   ||| |||::: ||| |||::: ||| |||::: |||
35 LeuAlaAlaLeuThrAsnGlyAsnIleLeuPheGluAspTyrProGl 51
   ::::::::::::::::::::|
49 .....RAGAGCGAAATCCCAAGGTCATAGAGAGGCATACATCCAG 91
   |||::: |||::: |||::: |||::: |||
51 yLeuGlyLysThrLeuLeuAlaLysValPheAlaArgVal..... 64
   ::::::::::::::::::::|
92 TCATCAGAGACACTGATTAAGAAAGAAATTCCTTTTGGCTCAACATAACG 141
   |||::: |||::: |||::: |||::: |||
65 .....IleGlyAlaAspTyrArgVal 72
   ::::::::::::::::::::|
142 GGTATACCTTAAAGTTCTCCGGAAGGATATTATAGACCTCGTTAAGG 191
   ::::::::::::::::::::|
73 GlnPheThrProAspLeuLeuProSerAspIleIle..... 84
   ::::::::::::::::::::|
192 GGCATC.....GCGAGTGACCTGATAGATAATC...GGAAGCA 229
   |||::: |||::: |||::: |||::: |||
85 .GlyValLysIleTyrArgGlyAspArgPheGluPheValLysGlyProI 101
   ::::::::::::::::::::|
230 GTTACACGCAACCAATCTC.....CCCTC 255
   |||::: |||::: |||::: |||::: |||
101 lePheThrAsnValLeuLeuAlaAspGluIleAsnArgSerProProlys 117
   ::::::::::::::::::::|
256 CTCCTCGCTTAGCAGAGTAGAAGCACAAAGTTCAGAGAGATAGGAAGTAA 305
   ::::::::::::::::::::|
118 ThrGlnAlaLeuLeuGluAlaMetGluGluLysGlnIleThrValGl 134
   ::::::::::::::::::::|
306 GGAAGAGCTCTTCGAGCTTTCTCCAAAGGATTCGTGGTCCGACAGCTCG 355
   ::::::::::::::::::::|
134 uGlyGluThrPheSerLeuSerMetProPhePheValLeu.....Alat 149
   ::::::::::::::::::::|
356 CCTATGACCGGATATCCCTCCACTGCAAGACACGGTTATGATAT 405
   ::::::::::::::::::::|
149 hrGlnAsnProfile..... 153

```

```

SQ SEQUENCE 602 AA: 70219 MW: 326B1198D54DD43E CRC64;

alignment_scores:
  Quality: 119.00      Length: 157
  Ratio: 1.545         Gaps: 4
  Percent Similarity: 49.045 Percent Identity: 26.752

alignment_block:
US-09-886-400-3 x Q9V0M7 ..
Align seg 1/1 to: Q9V0M7 from: 1 to: 602

40 GAATATCCCAAGACCGAAATCCCAAGAGCTCATAGAGAGCCATACATCC 89
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
71 GluTyrIleLysAlaGluPheAsnArgTyrMetGluArgLysIle.... 85
90 AGTCATCGACGACCTGATTAAAGAAGAAATTCCTTTGGCTCAACATAA 139
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
86 .....LysLeuMetLysGluAspLeuLysLysAlaAspGlyLysL 99
140 CGGCTATACCTTAAAGTTCCTCCCGAAG..... 168
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 euArgAsnAlaIleGluPheMetIleLysTyrPheLysAspValTyrGlu 115
169 .....GATATTATAGACCTCGTTAAAGGGGGCAT 197
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
116 TyrTrpSerLysIleAsnGlyAspIleIleGlyArgPheLysGlnLeuG 132
198 CGCAGAGTACCTGATAGAGATATATCGAAGAGCTACACGACGCAATAC 247
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 nAspGluGlyPheValGluIleIleThrSerAlaAlaThrHisGlyTyrL 149
248 TCCCGCTCCCTCCCGCTTAGCAGA...GTAGAAGCACAAAGTTCAGAGAT 294
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
149 euProLeuLeuGlyArgAspGluAlaIleAspAlaGlnIleLeuThrGly 165
295 AGGAAGATTAGGAAGAGCTCTTCGAGCTTCTCCAAAGGATTCGGCT 344
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
166 IleArgValTyrGluLysTyrPheGlyLysLysProArgLysIleTrpLe 182
345 GCCAGAGCTCGCTATGACCGGATA..... 369
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
182 uProGluCysAlaTyrArgProAspGlyLeuTrpLysSerProSerThrG 199
370 .....ATCGCTGCCCATACTAGAGAGCAACCGT 396
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
199 LyGluIleLysTrpArgLysGlyIleGluHisPheLeuLysLysTyrGly 215
397 TATGAGTATCTATTCCGCGAC 417
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
216 LeuGluPhePheValGlu 222

seq_name: sp_bacteriap:083377
seq_documentation_block:
ID_083377 PRELIMINARY; PRT; 526 AA.
AC 083377;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN TP0358.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,

406 CTATTCGCCGACGGGAGGCGATGCTTTTCTCAGCTCATCTCAACTCGC 455
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154 ...GluGlnGluGlyThrTyrProLeuProGluAlaGlnMetAspArgPh 169
456 GATAAGCCCAATTAAGCGCTCTATCACACCTTATAAAGGCCCAAGGG 505
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
169 eMetLeuArgMetArgProGlyTyrProGluSerIleGluGluMetG 186
506 AAAAGCGCTTTAGTACATCAGCTAT..... 531
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
186 LuIleLeuArgArgArgIleSerTyrPheArgLysAspProThrGluAsp 202
532 .....CTCCTTGCTCTCAGGAGCTTAGGAGCGAT 563
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
203 ValGluProValValSerLeuGluThrPheArgArgIleGlnAspAla 219
564 AAAGCTGCTTTTGAAGTAAAGTAAAGCGCTAAAGCGAGTCAAGACATCG 613
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
219 lGluAlaValTyrValAspLysSerIleLeuLysTyrIleSerGluLeu 236
614 AAGCGGTACCGCTTGGTGGCGCTGTAACACGGCTGATATGCTCGGCATC 663
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
236 alArgAla.....ThrArgGluHisGluLeuValGluLeuGlySer 249
664 .....GGAAGGCTCTCTTATGAACTCTAAGAAAGTGGCGAGCTG 704
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 SerProArgGlyGlyLeuAlaLeuLysLeuAlaArgAlaLeuAla 266
705 GATAGAGACAAAGCAACATCTT.....CTATCGGACCCGATA 745
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
266 lMetAspGlyArgAspPheValIleProAspAspValLysArgValAla 283
746 TAGAGCTTATGCTATAGGACATTCAGGCTACAGATGAGTGTGAG 795
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
283 alGluAlaLeuAlaHisArgValIleLeuLysPheGluTyrAlaValGlu 299
796 GGATTATTA.....GAGGTATAGACAGCTCAACTCGGAAGTGCCT 839
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 GlyLeuArgAlaGluValGluValGluGluIleLeuAsnSerValArgVa 316
840 TCCCTCAGAGCTGAAGCACAGTGAAGGGAG 870
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
316 lPro.....LysTyrGluAlaGlnGlu 323

seq_name: sp_archaeap:Q9V0M7
seq_documentation_block:
ID_Q9V0M7 PRELIMINARY; PRT; 602 AA.
AC Q9V0M7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 70.2 KDA PROTEIN.
GN PAB1857.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=25292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Hellig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AJ248285; CAB49676.1; -
DR InterPro; IPR001395; Aldo_ket_red.
DR InterPro; IPR003803; DUF200.
DR InterPro; IPR003583; HH1.1.
DR Pfam; PF02651; DUF200; 1.
DR SMART; SM00278; HH1; 1.
DR PROSITE; PS00062; ALDOKETC_REDUCTASE_2; UNKNOWN_1.
DR Hypothetical protein; Complete proteome.
KW

```

RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 RT Spirochete";
 RL Science 281:375-388(1998).
 DR EMBL; AE001215; AAC65344.1; -;
 DR TIGR; TP0358; -;
 DR InterPro; IPR003603; DUF200.
 DR Pfam; PF02651; DUF200; 1.
 KW Complete proteome.
 SQ SEQUENCE 526 AA; 59699 MW; D5F0C32CADB6C41A CRC64;

alignment_scores:

Quality: 111.00 Length: 216
 Ratio: 1.047 Gaps: 7
 Percent Similarity: 49.074 Percent Identity: 19.907

alignment_block:

US-09-886-400-3 x O83377 ..

Align seg 1/1 to: O83377 from: 1 to: 526

```

7  GCCTCGCTCTTCCACGGCAACCTCCAGTATGCCGAA..... 42
|||||:|||||:  |||||:  |||||:
10 AlaPheValLeuAspCysAsnLeuProPheValArgGlyAlaGlyAlase 26
43 .....ATCCCAAGAGCGGAATCCCAAGGTGCATAGAGAGGCATACA 85
|||||:|||||:  |||||:  |||||:
26 rSerLeuLeuAlaGluSerArgPhePheLeuGluLeuSerTyrThrTyrL 43
86 TCCAGCTCATC.....GAGACACTGATTAAAGAAGAAATCCCTTT 126
|||||:|||||:  |||||:  |||||:
43 euProLeuLeuArgLeuCysGluThrLeuGluArgGluValProPhe 59
127 GGGCTCAACATA..... 138
|||||:|||||:
60 AsnIleSerLeuAlaIleGlyProValLeuCysGluMetLeuAlaAsnAr 76
138 ..... 138
76 gValLeuMetAspArgTyrArgAlaLeuAspAlaLeuIleGluPheG 93
138 ..... 138
93 lyGluArgGluAlaIleArgLeuArgAsnSerLeuGlnGluArgValGln 109
139 ACGGCTATACCTTAAGTTCCTC.....CCGAAGGATATTATAGACT 182
|||:|||||:  |||||:  |||||:
110 AlaGluAlaValLeuArgSerLeuArgSerHisArgAspTyrPheAspHi 126
183 CGTTAAAGGGGGCATCGCGAGTGACCTG..... 210
|||||:|||||:  |||||:
126 sCysAspGlyAlaLeuLeuGluArgIleAsnHisPhePheArgThrGlys 143
211 ..ATAGAGATAATCGGAACGAGCTACACGCGCAATACTCCCTC... 255
|||||:|||||:  |||||:  |||||:
143 erileGluLeuLeuAlaThrAlaValAsnCysPheLeuProPheTyr 159
256 .....CTCCCGCTTAGCAGAGTGAAGCACCAAGTTTCAGAGATAGGGA 299
|||||:|||||:  |||||:
160 GlnAspMetProGluSer...IleSerAlaGluIle**MetGlyLeuL 175
300 AGTTAAGGAAGAGCTTCGAGCTTTCTCCAAAGGATTCTGGTCCGAG 349
|||||:|||||:  |||||:  |||||:
175 eAsnTyrArgLysHisPheSerIleProArgGlyPheTyrLeuProG 192
350 AGCTCGCTATGACCGCATATCCCTGCCCATCTGAAGGACAAACGGTAT 399
|||||:|||||:  |||||:  |||||:
192 luLeuGlyTyrAlaProAlaLeuGluArgThrIleLysSerTyrGlyPhe 208

```

```

400 GAGTATCTATTCCGCGACGGGAGCGGATGCTTTTCTCAGCTCATCTC 447
|||||:|||||:  |||||:  |||||:
209 SerTyrThrIleLeuGluThrHisSerPheLeuPheGlyThrArgVal 224

```

OM of: US-09-886-400-3 to: PIR_71.* out_format : pfs

Date: Jun 11, 2002 10:38 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame-n2p.model -DEV=xlh
-Q/cgn2_1/USPTO_spool/US09886400/runat_11063002_150914_17315/app_query.fasta.1.1169
-DB=PIR_71 -OFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-OGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=glosom62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR_MIN=100 -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09886400 -GCN1_1.121
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-886-400-3
Query length: 1095
Database: PIR_71.*
Database sequences: 283138
Database length: 96089334
Search time (sec): 40.510000
```

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
pir2:E71144	+ 1494.00	2201.43	4.3e-115	364	hypothetical protein PH0368 - P
pir2:AC2112	+ 169.00	236.91	7.9e-06	529	hypothetical protein alr2450 [A
pir2:S76831	+ 162.50	227.29	2.7e-05	529	hypothetical protein - Synchro
pir2:G71241	+ 161.50	224.04	3.4e-05	633	probable alpha-amylase - Pyroc
pir1:ALDYAT	+ 157.50	217.32	7.5e-05	686	amylase A (EC 3.2.1.-) - Dicty
pir2:F79197	+ 147.00	204.39	0.0005	527	uncharacterized conserved prote
pir2:E75206	+ 145.50	200.02	0.0007	655	alpha-amylase (or 4-alpha-gluc
pir2:B64501	+ 141.50	197.44	0.0014	467	alpha-amylase (EC 3.2.1.1) - Me
pir2:A49512	+ 141.50	194.19	0.0015	649	alpha-amylase (EC 3.2.1.1) - Py
pir2:D71011	+ 130.50	179.37	0.0120	560	hypothetical protein PH186 - P
pir2:G83826	+ 128.00	170.74	0.0220	923	hypothetical protein BH1415 [im
pir2:A75207	+ 124.00	160.97	0.0522	1362	amylomullulanase PAB0122 - Pyr
pir2:AC1970	+ 123.50	166.21	0.0488	744	hypothetical protein alr1310 [A
pir1:B69553	+ 120.50	169.98	0.0691	324	methanol dehydrogenase regulat
pir2:C75120	+ 119.00	161.64	0.1084	602	hypothetical protein PAB1857 -
pir2:S64314	+ 115.50	157.33	0.2056	551	probable membrane protein YGR02
pir2:AF1930	+ 113.00	147.09	0.3945	1069	hypothetical protein all0993 [
pir2:D140383	+ 112.00	137.77	0.3433	312	malate dehydrogenase (EC 1.1.1.
pir2:D71334	+ 111.00	151.13	0.4771	526	conserved hypothetical protein
pir2:T37544	+ 110.50	150.21	0.5273	536	hypothetical serine-rich protei
pir2:E90250	+ 109.50	150.52	0.6073	447	alpha amylase [imported] - Sul
pir2:T39903	+ 108.00	146.54	0.8466	534	serine-rich protein - fission y
pir2:B70859	+ 106.00	143.73	1.23	526	hypothetical protein RV3031 - X
pir2:E90270	+ 105.50	137.67	1.57	902	conserved hypothetical protein
pir2:G96587	+ 104.00	137.89	1.95	704	hypothetical protein T22422.2 [
pir2:F84044	+ 103.50	145.13	1.73	314	malate dehydrogenase cith [imp
pir2:S63131	+ 103.50	138.16	2.08	636	probable membrane protein YNL17
pir2:F70475	+ 101.50	134.18	3.13	705	VacB protein (ribonuclease II 4
pir2:B71541	+ 101.00	138.68	2.99	415	probable amino acid (glutamate)
pir1:S62421	+ 101.00	139.35	3.85	1067	endopeptidase La homolog (EC 3
pir2:G96943	+ 100.50	142.33	2.92	266	probable xylanase/chitin deacet
pir2:D71361	+ 100.50	133.99	3.66	619	probable alpha-amylase 1 - syph
pir2:T41144	+ 99.50	136.55	5.19	1131	hypothetical serine rich prote
pir2:T31067	+ 99.00	111.44	8.42	4845	BIR repeat containing ubiquitin
pir2:AE1926	+ 98.50	131.53	5.27	588	two-component response regulato
pir2:T19319	+ 98.00	132.34	5.56	503	hypothetical protein cl5H11.5 -
pir1:H70414	+ 98.00	126.21	6.22	764	conserved hypothetical protein
pir2:T46707	+ 97.50	134.29	5.68	383	proteophosphoglycan, membrane-a
pir2:T14710	+ 97.50	133.81	5.76	402	probable transposase - Yersinia
pir2:AB0193	+ 97.50	133.81	5.76	402	transposase, IS285 insertion el

```
pir2:AB0358 + 97.50 133.81 5.76 402 | transposase, IS285 insertio
pir2:AB0472 + 97.50 133.81 5.76 402 | transposase, IS285 insertio
pir2:AC0341 + 97.50 133.81 5.76 402 | transposase, IS285 insertio
pir2:AD0002 + 97.50 133.81 5.76 402 | transposase, IS285 insertio
pir2:AE0190 + 97.50 133.81 5.76 402 | transposase, IS285 insertio
```

seq_name: pir2:E71144

seq_documentation_block:

```
hypothetical protein PH0368 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 08-Sep-2000
C:Accession: E71144
R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; S
M.; Ohnuku, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Og
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil
A:Reference number: A71000; MUID:98344137
A:Accession: E71144
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-364 <RAW>
A:Cross-references: GB:AP000002; NID:g3236129; PIDN:BAA29442.1; PID:dl030385; PID:g3
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenB
C:Genetics:
A:Gene: PH0368
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0368
```

alignment_scores:

```
Quality: 1494.00 Length: 363
Ratio: 4.381 Gaps: 0
Percent Similarity: 93.939 Percent Identity: 75.758
alignment_block:
US-09-886-400-3 x E71144
Align seg 1/1 to: E71144 from: 1 to: 364
```

```
1 TTGAGAGGCTCGTCTTTTCACGGCAACCTCCAGTATCGCGAAATCCCAAA 50
: ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1 MetArgAlaLeuLeuPheHisGlyAsnLeuGlnTyAlaGluLeuProLy 17
51 GAGCGAAATCCCAAGGTCATAGAGAAGGCATATCCATCCATCATCGAGA 100
: ::::::::::::::::::::::::::::::::::::::::::::::::::::::
17 shisGluLeuSerLysValIleGluLysSerTyPheProThrIleSerG 34
101 CACTGATTAAGCAAGAAATCTCTTTGGGCTCAACATACGGGTATACC 150
: ::::::::::::::::::::::::::::::::::::::::::::::::::::::
34 luleuLeuLysArgGluLeuProPheGlyLeuAsnIleThrGlyTyrSer 50
151 TTAAAGTTCTCTCCGAGAGATATTATAGACCTCGTTAAAGGGGCATCGC 200
: ::::::::::::::::::::::::::::::::::::::::::::::::::::::
51 LeuGlnPheLeuProGlnGluLeuIleHisLeuLysGluGlyIleG1 67
201 GAGTACCTGATAGAGATAATCCGAGACGAGTACACGCGCATCTATCC 250
: ::::::::::::::::::::::::::::::::::::::::::::::::::::::
67 uSerGluLeuIleGluLeuIleGlyThrSerTyrThrHisAlaIleLeuP 84
251 CCCTCTCTCCGCTTAGCAGAGTAGAAGCACCAAGTTCAGAGAGATAGGAA 300
: ::::::::::::::::::::::::::::::::::::::::::::::::::::::
84 roLeuLeuThrLeuSerArgIleGluAlaGlnIleLysArgAspArgGlu 100
301 GTTAAGGAGAGACTCTTCGAGCTTTTCCAAAGGATTCCTGGCTGCCAGA 350
: ::::::::::::::::::::::::::::::::::::::::::::::::::::::
101 IleLysGluGluIlePheGluValSerProGlyGlyPheTrpLeuProG1 117
351 GCTCGCTATGACCGAGTAATCCCTGCCATATCTGAAGGACACCGTTATG 400
: ::::::::::::::::::::::::::::::::::::::::::::::::::::::
117 uLeuAlaTyraPpProIleIleProAlaIleLeuArgAspAsnGluTyrg 134
401 AGTATCTATTCGCCGAGGGGAGCGATGCTTTTCTCAGCTCATCTCAAC 450
```

```
|||||
134 IuTyLeuPheAlaAspGlyGluAlaMetLeuPheSerAsnHisLeuAsn 150
451 TCGGCGATAAGCAATTAACCGCTATCCACACCTTATTAAGGCCCA 500
|||||
151 SerAlaIleLysSerIleLysProLeuTyProTyLeuIleLysAlaG 167
501 AAGGAAAGCGCTTTAGGTACATACCTATCTCTGCTCAGGGAGC 550
|||||
167 nArgGlyGluGlyPheValTyLeuAsnTyLeuLeuGlyLeuArgGlu 184
551 TTAGGAAGCGGATAAAGCTGCTTTTGAAGCTAAGCTAAGCTAAGGCA 600
|||||
184 euLysLysAlaIleAsnLeuThPheGlyGlyLysValThrLeuGluAla 200
601 GTCAAGACATCGAAGCGGTACCGGTTTGGTGGCGGTGAACAGCGCT 650
|||||
201 ValLysAspIleGluAlaIleProValTrpValSerIleAsnIleAla 217
651 AATGCTGGCATCGAAGGCTTCTCTATGAATCCTTAAGAAAGTGCGCA 700
|||||
217 eMetLeuGlyAlaGlyArgPheProLeuMetSerProLysArgValAla 234
701 CTGGATAGAGGACAAAGCAACATCTTCTATACGGCCGATATAGAG 750
|||||
234 snTrpIleLysGlyLysAspGluIleLeuLeuTyGlyThrAspIleGlu 250
751 TTCATTGGCTATAGGACATTCAGGCTACAGAAATGAGTGTGAGGAT 800
|||||
251 PheLeuGlyTyArgSerIleAlaGlyHisLysIleAlaIleSerGlyLe 267
801 ATTAGAGTTATAGAGAGCTAACTCGGAAGCTGCTCTCCCTCCCTCAG 850
|||||
267 uMetGluValLeuAsnGluLeuGlyGlyLeuLeuCysLeuProArgAsp 284
851 TGAAGCAGTGGGAAGGAGCTCTACTACGAGCTTCGAGTGGGACCA 900
|||||
294 leArgHisAsnGlyArgLeuTyLeuArgThrSerTrpAlaPro 300
901 GATAAGCTTTGAGGATATGAGAGAGGAGGAGGAGGAGGAGGAGGAG 950
|||||
301 AspLysSerLeuArgIleTrpLysGluAspGluGlyAsnAlaArgLeu 317
951 TATGCTGCTACAAATATAGGCGGAGCTCGCCCTTTAGCGGAGCA 1000
|||||
317 nMetLeuThTyrcysMetaspGlyGluPheAlaPheLeuAlaGluAs 334
1001 CGGATGCAAGGGGATGGAGCCCTCCCTCAGAGAGGCTGGATGCTTC 1050
|||||
334 erAspAlaArgGlyTrpGluProLeuProGluArgLeuAspAlaPhe 350
1051 CGGCGATATATACGATTTGAGGGGTGAAATGGGAA 1089
|||||
351 LysAlaIleTyLysTyTrpArgAsnGluAsnGlyLys 363

seq_name: pir2:AC2112
seq_documentation_block:
hypothetical protein alr2450 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #tax_change 11-Jan-2002
C:Accession: AC2112
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriquchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Status: preliminary
A:Accession: AC2112
A:Molecule type: DNA
A:Residues: 1-529 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB74149.1; PID:gl7131542; GSPDB:GN00179
```

```
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr2450
C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

alignment_scores:
Quality: 169.00 Length: 495
Ratio: 0.820 Gaps: 21
Percent Similarity: 41.616 Percent Identity: 21.010

alignment_block:
US-09-886-400-3 x AC2112
Align seg 1/1 to: AC2112 from: 1 to: 529

7 CGCTCCTCTTTACGGCACCTCCAGTATGCGGAAATCCCAAGAGCGA 56
|||||
7 AlaleuValleuHisAlaHisLeuProPheValArgHisProGluSerAs 23
57 A.....ATCCAAAGTCATAGAGAAGGCATACA 85
:|||||
23 pTyrvAlleuGluGluGluTrpLeuTyGluAlaIleThrGluThTyri 40
86 TCCCA.....GTCATCGAGACACTGATTAAAGAAATTCCTTT 126
|||||
40 leProLeuLeuLysValPheAspGlyLeuLysArgAspGlyIleAspPhe 56
127 GGGCTC..... 132
:
57 LysIleThrMetSerMetThrProProLeuValSerMetLeuArgAspPr 73
132 ..... 132
73 oLeuLeuGlnGluArgTyArgAlaHisLeuAlaGlnLeuGluGluLeu 90
133 .....AACATACGGGCTATACCTTAAAGTTCCCTC 162
|||||
90 leGluLeuGluGlyGluArgAsnAlaGlnAsnGlyHisLeuArgTyLeu 106
163 CCGAAG..... 168
:
107 AlaGluHisTyAlaThrGluPheAsnGluAlaArgGlnMetTrpGluAr 123
169 .....GATATTATAGACCTCTTAAAGGGGCATCGCGAGTGACC 208
|||||
123 gTyvAsnGlyAspLeuValThrAlaPheLysGlnPheGlnAspSerAsn 140
209 TGATAGAGATAATCGGAACGAGCTACACGACCAATACCTCCCTCCCTC 258
|||||
140 snLeuGluIleIleThrCysGlyAlaThrHisGlyTyLeuProLeuMet 156
259 CCGCTT.....AGCAGATGAGACACAGTTTCAGAGATAGGGAAGT 302
:
157 LysMetTyvProGlnAlaValTrpAlaGlnIleGlnValAlaCysGluHi 173
303 TAAGGAAGAGCTCTTCGAGCTTTCTCCAAAGGATTCCTGCTGCCAGAGC 352
|||||
173 sTyvGluGluThrPheGlyArgProProLysGlyIleTrpLeuProGluC 190
353 TCGCTATGACCGGATTAATCCCTGCCATACCTAGGACACAGGTTATGAG 402
|||||
190 ysAlaTyTyvGluGlyLeuGluArgMetLeuAlaAspAlaGlyLeuArg 206
403 TATCTATTCCGCGAGGGGAGGATGCTTTTCTCAGCTCATCTCAACTC 452
|||||
207 TyvPheLeuThrAspGlyHisGlyIleLeuTy.....AlaArgProArgPr 222
453 GCGGATAAAGCCAAATTAAACCGCTC..... 477
:
222 oArgPheGlyThTyvAlaProIlePheThrGluThThrGlyValAlaAlaP 239
477 ..... 477
```


175 heProIleAspGluLysLeuArgTyrLeuIleProPheArgProValAsn 191
556 AAGCGATAAAGCTCGTGTTTTGAAGCTAAGGT 587
192 GluThrLeuGluTyrLeuHisSerLeuAlaAspGluSerLysVal 208
588 AACGCTAAAGCGAGTCAAAGACATCGAAGCCGTACCGTTGGGTGGCCG 637
208 lAlaVal...PheHisAspAspGlyGluLysPheGlyAlaTrpProGlyT 224
638 TGAACACGGCTGTAATGCTCGGCATCGGAAGCTTCCTTATGAATCCT 687
224 hrHisGluLeuValTyr..... 229
688 AAGAAAGTGGCGAGCTGGATA.....GA 710
230GluArgGlyTrpLeuLysGluPhePheAspArgIleSerSerAs 244
711 GGACAAGGACAACAATTCTTATATACGCACCGATATAGAGTTCATTGC. 759
244 pAspLysIleAsnLeuMetLeuTyrSer.....GluTyrLeuSerL 258
760TATAGGACATTCAGAGCTAC...AGA 783
258 ysPheArgProLysGlyLeuValTyrLeuProIleAlaSerTyrPheGlu 274
784 ATGAGTGTGAGGGATTA.....TTAGAGGT 809
275 MetSerGluTrpSerLeuProAlaArgGlnAlaLysLeuPhePheGluPh 291
810 TATAGCAGAGCTCAACTCGGAACGTGCCCTCCCTCAGAGCTGAAGCACA 859
291 eIleLysLysLeu...LysGluLeuAsnLeuPheGluLysTyrArg... 305
860 GTGGAAGGAGCTCTACTTACGGACTTCGAGTTGGCCACAGATAAAGAC 909
306IlePheValArgGlyGlyIleTrp.....LysAsn 315
910 TTGAGGATATGGACAGAGGAGGAGGAAC.....GCAACACATTA 950
316 Phe...LeuTyrLysTyrProGluLysTyrMetHisLysArgMetLe 331
951 TATGCTCTGCTTACAATATGAGGGCGCAACTGCCTTTTAGCCGAGAACA 1000
331 uMetLeuSer.....LysLeuLeuArgAsnAsn. 340
1001 GCGATGAAGGGGATGGAGGCCCTCCCTGAGAGGAGCTGGATGCCTTC 1050
341ProThrAlaArgIlePheValLeu 348
1051 CGGCGCATATATACGAT.....TGGAGGGGTGAAATGGG 1086
349 ArgAlaGlnCysAsnAspAlaTyrTrpHisGlyValPheGly 362

seq_name: p1r1:ALDYAT

seq_documentation_block:
 amylase A (EC 3.2.1.-) - Dictyoglomus thermophilum
 C:Species: Dictyoglomus thermophilum
 C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 18-Jun-1999
 C:Accession: S00628; A34969
 E:Fukusumi, S.; Kamizono, A.; Horinouchi, S.; Beppu, T.
 Eur. J. Biochem. 174, 13-28, 1988
 A:Title: Cloning and nucleotide sequence of a heat-stable amylase gene from an anaerobic
 A:Reference number: S00628; MUID:88225097
 A:Accession: S00628
 A:Molecule type: DNA
 A:Residues: 1-686 <FUK>
 A:Cross-references: EMBL:X07896; NID:g2688; PIDN:CMA30735.1; PID:g2689
 A:Accession: A34969
 A:Molecule type: protein
 A:Residues: 2-13 <FUK2>
 C:Genetics:


```

200 sPLysPheAlaSerGluAspLysSerLysLysLeuLeuPheAsp 216
|||||
745 ATAGAGTTTCATGGC.....TATAGGCACATTGCCAGGTA 779
|||||
217 GlyLysPheGlyLeuTyrProAspThrTyrArgThrVal.....Ty 231
|||||
780 CAGATAGTGTTCAGGGATTATAGGTT.....ATAGACGAGCTCACT 826
|||||
231 rGlu.....GluGlyTrpLeuGluThrPheValSerLysIleLysG 245
|||||
827 CGGAACGTGCTCCCTCA.....GAG 849
|||||
245 LuAsnPheLeuValThrProValAsnLeuTyrThrTyrMetGlnArg 261
|||||
850 CTGAAGCACAGTGGAGGGAGCTCTACTTACGAGCTTCGAGTTGG 894
|||||
262 ValLysProLysGlyArg.....IleTyrLeuProThrAlaSerTyr 275
|||||
seq_name: pir2:F97197

```

```

seq_documentation_block:
uncharacterized conserved protein, related to alpha-amylase/alpha-mannosidase CAC2414 [1]
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: F97197
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4833-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97197
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80369.1; PID:g15025430; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

```

```

alignment_scores:
Quality: 147.00 Length: 486
Ratio: 0.766 Gaps: 20
Percent Similarity: 39.506 Percent Identity: 18.930

```

```

alignment_block:
US-09-886-400-3 x F97197 ..

```

Align seg 1/1 to: F97197 from: 1 to: 527

```

13 GTCTTTCACGGCAACCTCCAGTATGCCGAATCCCAAGAGCGAA..... 57
|||||
9 ValLeuHisSerHisMetProPheValArgHisProGluThrArgAspSe 25
58 .....ATCCCAAGTCCATAGAGGCGCATACATCCCGAG 91
|||||
25 rLeuGluLysArgTrpLeuPheGluAlaMetSerGluCysTyrIleProL 42
|||||
92 TCATCCGAG.....ACACTGATTAAAGAAATTCCTTTTGGG... 129
|||||
42 euLleGluValTyrAspAsnLeuLeuLysAspAsnIleLysPheArgMet 58
|||||
129 ..... 129
59 ThrMetSerIleThrProProLeuMetSerMetLeuGlnAspGluTyrLe 75
130 .....CTCAACATAACGGCGGTATACCTTAAAGTTCCTCCCGGA 166
|||||
75 uAsnSerArgTyrLeuAsnTyrLeuLysLysThrIleGluLeuSerGluL 92
|||||
167 AGGTATTATA..... 177

```

```

|||||
92 ysGluIleLeuArgThrLysAsnAsnArgGluGluAsnLysValAlaLeu 108
177 ..... 177
109 PheTyrAsnLysArgAlaGluAsnThrLeuLysIleTyrGluLysTyrAs 125
178 ....GACCTCGTTAAAGGGGCATCCGAGTGAC.....CTGATAG 214
|||||
125 pAsnAsnLeuIleAsnAlaPheArgLysTyrAspArgLeuGlyCysValG 142
215 AGATATATCGAAGAGTACACGACGCAATACTCCCTCCCTCCCGGTT 264
|||||
142 luLleIleThrCysAlaIaThrHisAlaLeuLeuProLeuIleLeuIle 158
265 AGCAGA.....GTAGAAGCACAAAGTTTCAGAGATAGGGAAGTTAAAGA 308
|||||
159 AsnArgGlnAlaValLysAlaGlnIleAlaThrGlyValGlnSerTyrI 175
309 AGAGCTTTCGAGCTTCTCCRAAGGATTCTGGTCCGACAGCTCGCT 358
175 eaSnThrMetGlyHisGluProAsnGlyIleTrpLeuProGluCysAla 192
359 ATGACCCGATAATCCCTGCCATCTACTAGGACACACGGTTATGATATCTA 408
192 yrThrTyrGlyIleAspAsnIleLeuSerGluPheGlyIleLysTyrPhe 208
409 TTGCCCGACGGGAGCGGATGCTTTTCTCAGCTCATCTCACTCGCGCAT 458
209 IleSerGluGlyLysAlaIleAspTyrAlaSerPro..... 220
459 AAAGCCAATTAACCCGCTCTATCCACACCTTATAAAGGCCCAAGGAAA 508
221 .....LysProMetTyr..... 224
509 AGCGCTTATAGTACATCAGCTATCTCTGGTCTCAGGAGCTTAGGAAG 558
224 ..... 224
559 GCGATAAAGCTCGTTTTTGAAGGTAAGGTAACGCTAAAGGCAGTCAAAGA 608
224 ..... 224
609 CATCGAAGCGGTACCCGTTTGGTGGCCGTGAACACGGCTGTAATGCTCG 658
225 .....GlyThrAsnThr..... 228
659 GCATCGAAGGCTTCCTCTATGATCTTAAGAAAGTGCGAGCTGGATA 708
229 .....ProIleAlaAlaProSerGlyValCysAlaPheGly 240
709 GAGGACAAGGACAACATCTCTATACGCGCACCGATATAGAGTTTCATTGG 758
241 ArgAspMetAspSer.....SerTyrGlnValTrpSerAspPheMetG1 255
759 C.....TATAGGACATTCGAGGCT 778
255 YTYrProGlyAspPheAsnTyrArgGluPheTyrArgAspIle...GlyP 271
779 ACAGAATGAGTGTAGGAGTATTAGAGGTTATAGACGAGCTCAACTCG 828
271 heGluLeuProMetGluTyrIleLysProTyrIleAsnGlu.....Asn 285
829 GAAGTGGCTTCCTCCAGAGCTGAAGCAC.....AG 860
286 GlyIleArgIleAspThrGlyPheLysTyrTyrLysIleThrGlyAsnSe 302
861 TGAAGGAGAG...CTCTACTTACGGACTTCGAGT.....TGGG 895
302 rGlyGluLysGlyIleTyrAsnArgGluAsnAlaMetLysLysValTpg 319
896 CACCAGATAAGAGCTTGAGGATATGGAGAGAGGACGAGGGGAC...GCA 942

```

```
319 LUHisAlaSerHisPheAlaSerCysArgHisAspGlnIleAsnAlaAla 335
943 AGACITATATG ..... 954
336 AlaAlaAsnMetAspLysProProIleIleThrCysProTyrAspThrG1 352
954 ..... 954
352 uLeuTyrGlyHisTrpTrpPheGluGlyProAspPheIleAsnAlaPheI 369
955 .....CTGTCTACATATGAGGGCGAATC 981
369 leArgLysSerAlaGluAspTrpThrSerTyrGluLeuIleThrProThr 385
982 GCCCTTTTACCGCAGACAGCATGACAGGGGATGGAGCCCTCCCTGA 1031
386 GluTyrLeuLysAsnAsnSerMetValGlnCysSerProSerPro... 401
1032 GAGGAGGCTGGATGCTTCCGGCGCATATATACCATGGAGGGGTGAAA 1081
402 .....SerSerTrp...GlyGluA 407
1082 ATGGGGAA 1089
407 snGlyAsp 409
seq_name: pir2:E75206
seq_documentation_block:
alpha-amyase (or 4-alpha-glucanotransferase) PAB0118 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: E75206
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: E75206
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-655 <KAW>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:95457433; PIDN:CAB49100.1; PID:9545760
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: amYA; PAB0118
C:Superfamily: Dictyoglomus thermophilum amylase A
alignment_scores:
Quality: 145.50 Length: 401
Ratio: 0.808 Gaps: 22
Percent Similarity: 44.888 Percent identity: 22.444
alignment_block:
US-09-886-400-3 x E75206
Align seg 1/1 to: E75206 from: 1 to: 655
73 GAGAGGCATACATCCAGCTACGACACTGATTAAGAAGAATTC 122
|||||
28 GluLysAlaTyrArgProPheLeuGluIleLeu.....GluGluTyrPr 42
123 T.....TTGGGCTCAACATACCGGCTATACCTTAAGTTTCPC 163
|.....|
42 oAsnMetLysValAlaIleHisIleSerGlyIleLeuValGluTrpLeuG 59
164 CGAAG.....GATATATAGACCTCGTTAAAGGGGATCCGAGT 204
...|
59 luGluAsnLysProAspTyrIleAspLeuLysSerLeuValArgLys 75
205 GACCTGATAGATATACGACGAGCTACACGACGAATACTCCCTC 254
|||||
76 GlyGlnValGluIleValAlaGlyPheTyrGluProValLeuAlaI 92
```

```
255 CCTCCGCTTAGCAGTAGAAGCACAAAGTTTCAGAGAGATAGGGAAGTTA 304
|||||
92 aileProLysGluAspArgLeuGluGlnIleTyrLeuLeuLysGluTrpA 109
305 AGAAGAGCTCTTCGAGCTTTCOCAAAAGGATTCTGGCTCCAGAGCTC 354
|||||
109 laLysLysIle...GlyTyrAspAlaLysGlyLeuTrpLeuThrGluArg 124
355 GCCTATGACCCGATATCCCTCCCATCTAGAGGACACACGCTTATGAT 404
|||||
125 ValTrpGlnProGluLeuValLysThrLeuArgGluAlaGlyIleGluTy 141
405 TCTATTCGCCGACGGGAGGGATGCTTTTCTCAGCTCATCTCAACTCGG 454
|||||
141 rValValValAsp.....AspTyrHisPheMetSerA 152
455 CG...ATAAGCCCAATTAAACCGCTCTATCCACACCTTATAAGGCCCA 501
||
152 laGlyLeuSerLysAspGlnLeuPheTrpProTyrTyrThrGluAspGly 168
502 AGGAA.....AAGCGCTTTAGGTAC..... 522
||
169 GlyGluValIleThrValPheProIleAspGluLysLeuArgTyrLeuI 185
523 .....ATCAGCTATCTCCTTGTCTCAGG 547
|||||
185 eProPheArgProValAspLysValIleSerTyrLeuHisSerLeuAlaS 202
548 AGCTTAGGAAGCGGATAAGCTGCTTTTGAAGGTAAAGTAACTAAAG 597
||
202 erGluAspGluSerLysValAlaValPheHis..... 212
598 GCAGTCAAGACATCGAAGCCGTACCCCTTTGGGTGGCGCTGACACGGC 647
||
213 .....AspAspGlyGluLysPheGlyIleTrp..... 221
648 TGTAATGCTCGGCATCGGAAGCTTCTCTTATCAATCCTAAGAAAGTGG 697
||
222 .....ProMet.....Thr 225
698 CGAGCTGGATAGGACACAGGAC..... 720
||
225 yrGluTrpValTyrGluLysGlyTyrLeuArgGluPhePheAspArgVal 241
721 .....AACATTCTCTATACGGCACCGCATATAGAG... 750
|||||
242 SerSerAspLysAlaIleAsnIleMetLeuTyrSerGluTyrLeuGlnLy 258
751 .....TTCATTGGCTATAGGACATTCGACGGCTAC..... 780
||
258 sPheLysProLysGlyLeuValTyrLeuProIleAlaSerTyrPheGluM 275
781 .....AGAATGAGTGTGAGGA 798
275 etSerGluTrpSerLeuProAlaGlnGlnAlaLysLeuPheValGlu... 290
799 TTATTAGAGTTTATAGACGAGCTCAACTCGGAAGTGTGCTTCCCTCAGA 848
|||||
291 PheValGluLysLeuLysGluLeuAsn..... 299
849 GCTGAAGCAGACATGGAAGGACCTCTACTTACGGACTTCGAGTTGGGCAC 898
||
300 ...MetPheGluArgTyrArgValPheValArgGlyGlyIleTrp.... 313
899 CAGATAAGAGCTTGGAGTATATGGAGAGGAGGAGGGAAC..... 939
|||||
314 .....LysAsnPhe...PheTyrLysTyrProGluAlaAsnTyrMetHis 327
940 GCAAGACTTTAATATGCTGTCTACAAATATGAGGGCGCAACTCGCCCTTT 989
|||||
328 LysArgMetLeuMetLeuSer.....ArgLeuLe 337
```


C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
 C:Accession: A49512
 R:Lademan, K.A.; Asada, K.; Uemori, T.; Mukai, H.; Taguchi, Y.; Kato, I.; Anfinsen, C.E.
 J. Biol. Chem. 268, 24402-24407, 1993
 A:Title: alpha-Amylase from the hyperthermophilic archaeobacterium Pyrococcus furiosus. C
 A:Reference number: A49512; MUID:94043280
 A:Accession: A49512
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-649 <LAD>
 A:Cross-references: GB:L22346; NID:g347939; PIDN:AAAT2035.1; PID:g347940
 C:Genetics:
 A:Gene: amyA
 A:Start codon: GTG
 C:Function:
 A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
 A:Pathway: glycogen/starch degradation
 C:Superfamily: Dictyoglomus thermophilum amylase A
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

alignment_scores:
 Quality: 141.50 Length: 391
 Ratio: 0.773 Gaps: 23
 Percent Similarity: 46.803 Percent Identity: 22.762

alignment_block:
 US-09-886-400-3 x A49512 ..
 Align seg 1/1 to: A49512 from: 1 to: 649
 73 GAGAGGCATACATCCAGTCATCGACACATGATTAAAGAAATTC 122
 |||||:||||| ||| :||||| ||||| ||||| |||||
 29 GluLysCysTyrTrpProPheLeuGluThrLeu.....GluGluTyrPr 43
 123 T.....TTTGGCCACATAAAGGGCTATACCTTAAAGTTCCTCC 163
 :|||:|||||: |||: |||||: |||||: |||||: |||||
 43 OasnMetLysValAlaIleHisThrSerGlyProLeuIleGluTrpLeuG 60
 164 CG.....AAGATATTATACCTCGTTAAAGGGGCAATCCGAGT 204
 :|||: |||||: |||||: |||||: |||||: |||||: |||||
 60 InAspAsnArgProGluTyrIleAspLeuLeuArgSerLeuValLysArg 76
 205 GACCTGTAGAGATAATCGGACGAGGTACACGACGCAATATCTCCCT 254
 :|||: |||||: |||||: |||||: |||||: |||||: |||||
 77 GlyGlnValGluIleValAlaGlyPheTyrGluProValLeuAla 93
 255 CTTCCCGCTTAGCAGAGTAGAAGCAAGTTCAGAGAGATAGGAAGTTA 304
 :|||: |||||: |||||: |||||: |||||: |||||: |||||
 93 rIlePro.....LysGluAspArgIleGluGlnIleArgLeuMetL 107
 305 AGGAA.....GAGCTCTCGAGCTTTCTCCAAAGGATTCGTGCTCCA 348
 |||||: |||||: |||||: |||||: |||||: |||||: |||||
 107 ysGluTrpAlaLysSerIleGlyPheAspAlaArgGlyValTrpLeuThr 123
 349 GAGCTCGGCTATGACCCGATAATCCCTGCCATCTGAAGACACGGTTA 398
 ||| :|||||: ||| :|||||: ||| :|||||: ||| :|||||
 124 GluArgValTrpGlnProGluLeuValLysThrLeuLysGluSerGlyI 140
 399 TGAGTATCTATTGCGCGACGGGAGCGATCTTTCTCAGCATCTCA 448
 :|||: |||||: |||||: |||||: |||||: |||||: |||||
 140 eAspTyrValIleValAsp.....AspTyrHisPheM 151
 449 ACTCGCGCATAAAGCAATTAACCGCTCTAT...CCACACTTATAAAG 495
 |||||: |||||: |||||: |||||: |||||: |||||: |||||
 151 etSerAlaGlyLeuSerLysGluGluLeuTyrTrpProTyrThrGlu 167
 496 GCCCAAGGGAA.....AAGCGCTTAGGTA 521
 |||||: |||||: |||||: |||||: |||||: |||||: |||||
 168 AspGlyGlyGluValIleAlaValPheProIleAspGluLysLeuArg.. 183
 522 CATCAGCTATCTCTTGTCTCAGGAGCTTAGGAAGCGCATAAAGCTC. 570
 |||||: |||||: |||||: |||||: |||||: |||||: |||||

184TyrLeuIleProPheArgProValAspLysValLeuGluTyrL 198
 571GTTTTCAAGGTAAAGTAACGCTAAAGGCATC.....AAA 606
 :|||: |||||: |||||: |||||: |||||: |||||: |||||
 198 euHisSerLeuIleAspGlyAspGluSerLysValAlaValPheHisAsp 214
 607 GACATCGAAGCCGTACCGTTTGGTGGCCGTGAACACACGGCTGTAATGCT 656
 |||||: |||||: |||||: |||||: |||||: |||||: |||||
 215 AspGlyGluLysPheGlyIleTrpProGlyThrTyrGluTrpValTyr.. 230
 657 CGGCATCGAAGGCTTCCTCTATGAATCCTTAAGAAAGTGGCGAGCTGGA 706
 231GluLysGlyTrpL 235
 707 TA.....GAGGACAAGGACACATCTT 729
 :|||: |||||: |||||: |||||: |||||: |||||: |||||
 235 euArgGluPhePheAspArgIleSerSerAspGluLysIleAsnLeuMet 251
 730 CTATAGCGCACCGATATAGAG.....TTCATTGGCTATAG 764
 |||||: |||||: |||||: |||||: |||||: |||||: |||||
 252 LeuTyrThrGluTyrLeuGluLysTyrLysProArgLysLeuValTyrLe 268
 765 GCACATTGCAGCTACAGATGAGTGTGGGATATTAGAGCTTATAG 814
 :|||: |||||: |||||: |||||: |||||: |||||: |||||
 268 uproIleAlaSerTyr.....PheGluMet.... 276
 815 ACGAGCTCACTCGGAAGTGCCTTCCC..... 843
 |||||: |||||: |||||: |||||: |||||: |||||: |||||
 277SerGluTrpSerLeuProAlaLysGlnAlaArgLeuPhe 289
 844TCAGAGCTGAAGCACAGTGA.....AGGGA 869
 :|||: |||||: |||||: |||||: |||||: |||||: |||||
 290 ValGluPheValAsnGluLeuLysValLysGlyIlePheGluLysTyrAr 306
 870 GCTCTACTACGACTTCGAGTTGGCCACCATAGATAGAGGTTAGGATAT 919
 :|||: |||||: |||||: |||||: |||||: |||||: |||||
 306 gValPheValArgGlyGlyIleTrp.....LysAsnPhe...PheT 319
 920 CGAGAGAGAGAGAGGGAAC.....GCAAGACTTAATATGCTGCTCC 960
 :|||: |||||: |||||: |||||: |||||: |||||: |||||
 319 YLysTyrProGluSerAsnTyrMetHisLysArgMetLeuMetValSer 335
 961 TACAATATGAGGGCGAACTCGCCCTTTTACCGGAGAACACGCGATGCAAG 1010
 :|||: |||||: |||||: |||||: |||||: |||||: |||||
 336 LysLeuValArgAsnAsn..... 341
 1011 GCGATGGAGGCCCTCCTCGAGAGAGGCTGGATCGCTTCGGGCGGATAT 1060
 :|||: |||||: |||||: |||||: |||||: |||||: |||||
 342ProGluAlaArgLysTyrLeuLeuArgAlaGlnC 353
 1061 ATAACGAT.....TGGAGGGGT 1077
 |||||: |||||: |||||: |||||: |||||: |||||: |||||
 353 ysAsnAspAlaTyrTrpHisGly 360

seq_name: pir2:D71011

seq_documentation_block:

hypothetical protein PH1386 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000

C:Accession: D71011

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hino, Y.; Yamamoto, S.; S

M.; Ofuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Og

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A:Reference number: A71000; MUID:98344137

A:Accession: D71011

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-560 <RAW>

A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30492.1; PID:g3257809

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenB.

C:Genetics:
A:Gene: PH1386
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

```
alignment_scores:
  Quality: 130.50      Length: 234
  Ratio: 1.186         Gaps: 12
  Percent Similarity: 47.009      Percent Identity: 25.641

alignment_block:
US-09-886-400-3 x D71011      ..
Align seg 1/1 to: D71011 from: 1 to: 560

40  GAAATCCCAAGACCGCAATCCCAAGGTCTATAGAGAAGGCATACATCCC 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71  GlutryrileysArgGluPheGluLysTyrMetGluArgLysLeuLysSe 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90  AGTCATCGAGACACTG.....ATTAAAGAGAA 118
   : : : : : : : : : : : : : : : : : : : : : : : : :
87  rMetGluGluAspGluArgPheLysaspGluLysLeuArgGluAlaI 104
   : : : : : : : : : : : : : : : : : : : : : : : : :
119  TTCCTTTGGGCTCAACATAACGGCTATACCTTAAAGTTCCTCCGAAG 168
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104  leAsnPhe.....MetileGlyTyrPhe.....Lys 112
   ||| : : : : : : : : : : : : : : : : : : : : : : :
169  GATATTATAGACCTC.....GTTAAGGGGGCATCGCGAGTACCT 209
   ||| : : : : : : : : : : : : : : : : : : : : : : :
113  AspValTyrSerTyrTrpLysSerileAspGlyAsnileLeuGlyLysPh 129
   ||| : : : : : : : : : : : : : : : : : : : : : : :
210  G.....ATAGAGATAATCGGAACGAGCTACA 235
   : : : : : : : : : : : : : : : : : : : : : : :
129  eArgGluLeuGlnAspGluGlyTyrValGluValIleThrSerAlaAlat 146
   : : : : : : : : : : : : : : : : : : : : : : :
236  CGCAGCAATACCTCCCTCCTCCCTTAGCAGA...GTAGNAGCACAA 282
   ||| : : : : : : : : : : : : : : : : : : : : : : :
146  hrHisGlyTyrLeuProleuLeuGlyArgAspGluAlaIleGluAlaGln 162
   ||| : : : : : : : : : : : : : : : : : : : : : : :
283  GTTCAGAGAGATAGGGAAGTAAAGGAGAGCTCTCGAGCTTTCTCCAAA 332
   ||| : : : : : : : : : : : : : : : : : : : : : : :
163  LeuLeuAsnGlyIleLysValTyrGluLysTyrPheGlyArgLysProar 179
   ||| : : : : : : : : : : : : : : : : : : : : : : :
333  GGGATTCTGGCTCCGAGACTCCCTATGACCGGATA..... 369
   ||| : : : : : : : : : : : : : : : : : : : : : : :
179  gGlyIleTrpLeuProGluCysAlaTyrArgProaspGlyLeuTrpLysS 196
   ||| : : : : : : : : : : : : : : : : : : : : : : :
370  .....ATCCCTGCCATACG 384
   ||| : : : : : : : : : : : : : : : : : : : : : : :
196  erProSerThrGlyGluValLysTrpArgLysGlyIleGluHisPheLeu 212
   ||| : : : : : : : : : : : : : : : : : : : : : : :
385  AAGCACAAACGGTTATAGTATCTATTTCGGCGAGGGGAGCGATGCTTTT 434
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
213  LysLysPheGlyIleGluTyrPhePheValGlu..... 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
435  CTCAGCTCATCTCAACTCGCGGATAAAGCAATTAACCGCTCTATCCAC 484
   : : : : : : : : : : : : : : : : : : : : : : :
224  ....SerHisLeuIleaspLys...GlyProValSerLeuArgTyrGlyA 238
   : : : : : : : : : : : : : : : : : : : : : : :
485  ACCTATAAGGCCCAAGGAAACGGCTTTAGGTACATCAGCTATCTC 534
   : : : : : : : : : : : : : : : : : : : : : : :
238  snIleLeuProAla...LysThrLysArgSerThrLeuArgProTyrPhe 253
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
535  CTTGGTCTCAGGAGCTTAGGAAGGGATAAGCTGTTTTCGAAGGTAA 584
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254  .....LeuLysAsnGlyIleAlaValPheAlaArgAsnAr 265
   : : : : : : : : : : : : : : : : : : : : : : :
585  GGTAAACGCTAAGCGAGTCAAGACATCGAAGCCGCTACCCGTTGGGTGG 634
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
265  gGluThr.....GlyIleGlnValTrpSerA 274
   : : : : : : : : : : : : : : : : : : : : : : :
635  CC 636
   |||
```

274 la 274

seq_name: pir2:G83826

seq_documentation_block:
hypothetical protein BH1415 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G83826
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: AB3650; MUID:20512582; PMID:11058132
A:Accession: G83826
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-923 <STO>
A:Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA05134.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1415

alignment_scores:
Quality: 128.00 Length: 482
Ratio: 0.612 Gaps: 24
Percent Similarity: 43.361 Percent Identity: 20.124

alignment_block:
US-09-886-400-3 x G83826 ..

Align seg 1/1 to: G83826 from: 1 to: 923

```
7  GCGCTGCTTTTACGCGCACTCCAGATGCGGAATCCCAAGAGCGA 56
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7  SerleuValleuHisAlaHisleuProTyrValArgHisGlnGluGluAs 23
   ||| : : : : : : : : : : : : : : : : : : : : : : :
57  A.....ATCCAAAGGTGATAGAGAAGGCATACATCC 88
   : : : : : : : : : : : : : : : : : : : : : : :
23  pArgLeuGluGluArgTrpLeuPheGluAlaMetSerGluThrTrpIle 40
   ||| : : : : : : : : : : : : : : : : : : : : : : :
89  CAGTCATCGAGACACTGATTAAGAGAAATTCCTTTGGGCTCAACATA 138
   ||| : : : : : : : : : : : : : : : : : : : : : : :
40  roleuLeuTrpAlaLeuGluLysLeuProValLysHisAlaValThrIle 56
   ||| : : : : : : : : : : : : : : : : : : : : : : :
139  .....AC 140
   ||| : : : : : : : : : : : : : : : : : : : : : : :
57  SerPheThrProValMetGluMetLeuSerAspProLeuValGlnTh 73
   ||| : : : : : : : : : : : : : : : : : : : : : : :
141  GGCTAT.....ACTTAAAGTTCCTCCCGAAGGAT. 171
   ||| ||| : : : : : : : : : : : : : : : : : : : : :
73  rArgTyrLeuAsnHisLeuGluAsnThrGluGlnLeuLysLysGluG 90
   ||| : : : : : : : : : : : : : : : : : : : : : : :
172  .....ATTATAGACCTCGTTAA 189
   : : : : : : : : : : : : : : : : : : : : : : :
90  LuLysArgThrAsnAspGlnArgThrGlnAsnLeuValGlnPheTyrLys 106
   ||| : : : : : : : : : : : : : : : : : : : : : : :
190  GGG..... 192
   ||| : : : : : : : : : : : : : : : : : : : : : : :
107  GlnArgTyrGluLysLeuLysAlaThrPheLeuGlnTrpAspArgAsnLe 123
   ||| : : : : : : : : : : : : : : : : : : : : : : :
193  .....GGCATCGCGAGTGACCTGATAGAG.....ATAA 220
   ||| : : : : : : : : : : : : : : : : : : : : : : :
123  uLeuIleGlyPheArgSer...LeuMetGluAsnGluGlnCysThrLeu 139
   ||| : : : : : : : : : : : : : : : : : : : : : : :
221  TCGAAGCAGGACTACAGCAGCAATACTCCCTCTCTCCGCTTAGCAGA 270
   : : : : : : : : : : : : : : : : : : : : : : :
139  etThrSerAlaAlaThrHisAlaPhePheProTyrLeuLysThrLysGlu 155
   : : : : : : : : : : : : : : : : : : : : : : :
271  ...GTAGAAGCACAACTTCAGAGAGATAGGAGTTAAGGAGAGACTCTT 317
   : : : : : : : : : : : : : : : : : : : : : : :
156  AlaIleArgAlaGlnValArgHisGlyIleAlaCysPheGluGlnHisPh 172
   ||| : : : : : : : : : : : : : : : : : : : : : : :
```



```
466 .....ATTAAACCGCTCTATCCACACCTTATAAGAGCCCAAGGAA 508
      : : : : : : : : : : : : : : : : : : : : : : : : :
393 rAspAlaValLysAsnPheValGluGluLeuLysIleGlnLysGlnA 410
      : : : : : : : : : : : : : : : : : : : : : : : : :
509 AGCGCTTAGGTACATCAGCTACTCTCTGCTGCTCAGGAGCTTAGGA 558
      : : : : : : : : : : : : : : : : : : : : : : : : :
410 snTyAspGlySerLeuValTyValIleThrLeuAspGlyGluAsnPro 426
      : : : : : : : : : : : : : : : : : : : : : : : : :
559 GCGATAAAGCTCTTTTGAAGTAAGTAAGCTAAAGGCAGTC...AA 605
      : : : : : : : : : : : : : : : : : : : : : : : : :
427 TrpGluHisTyProPheAspGlyLysLeuPheLeuGluGluLeuTyAr 443
      : : : : : : : : : : : : : : : : : : : : : : : : :
506 AGACATCGACCGCTACCCGTTTGGGTGGCGGTGAACAGCGCTGTA... 651
      : : : : : : : : : : : : : : : : : : : : : : : : :
443 gGlnLeuGluGluLeuGlnLysGlyLeuIleArgThrValThrProS 460
      : : : : : : : : : : : : : : : : : : : : : : : : :
552 .....ATGCTCGGC.....ATCGAAGGCTTCTCTCTATTG 681
      : : : : : : : : : : : : : : : : : : : : : : : : :
460 erGluTyIleGluMetPheGlyAspLysAlaAsnLysLeuThrProLys 476
      : : : : : : : : : : : : : : : : : : : : : : : : :
582 AATCCTAAGAAGTGGCGGAGCTGGATAGGACAGGACAACTTCTTCT 731
      : : : : : : : : : : : : : : : : : : : : : : : : :
477 MetMetLysArgLeuAspPheThrThrGluAspAsnValAsnAlaLeu 493
      : : : : : : : : : : : : : : : : : : : : : : : : :
732 ATACGGCACCGATATAGAGTTCATTGGCTATAGGACATTGCAGGCTACA 781
      : : : : : : : : : : : : : : : : : : : : : : : : :
493 uLysAlaLysThrLeu.....GlyGluL 501
      : : : : : : : : : : : : : : : : : : : : : : : : :
782 GAATGAGTGTGAGGAGNTATTAGAGGTTATAGACGAGTCACACTCGGAA 831
      : : : : : : : : : : : : : : : : : : : : : : : : :
501 euTyAspMetValGlyValThrGlu..... 509
      : : : : : : : : : : : : : : : : : : : : : : : : :
832 CTGTGCTCTCCCTCAGAGCTGAAGCACAGTGAAGGAGGCTCTACTTACG 881
      : : : : : : : : : : : : : : : : : : : : : : : : :
510 .....GluMetGlnTrpPr 514
      : : : : : : : : : : : : : : : : : : : : : : : : :
882 GACTTCGAGTGGGCACCATGATAGAGCTTGGAGATATGAGAGAGGAGC 931
      : : : : : : : : : : : : : : : : : : : : : : : : :
514 oGluSerSerTrpIle...AspGlyThrLeuSerThrTrp.....I 527
      : : : : : : : : : : : : : : : : : : : : : : : : :
932 AAGGAGCAGCAAGACTTAATGCTGTCTCTACATATGAGGGCGCACTC 981
      : : : : : : : : : : : : : : : : : : : : : : : : :
527 leGlyGluProGlnGluAsnIleAlaTrpTyTrpLeuTyLeuAlaArg 543
      : : : : : : : : : : : : : : : : : : : : : : : : :
982 GCCCTTTTAGCGAGAACAGCAT...GAAGGGGATGGGAGCCCTCCC 1028
      : : : : : : : : : : : : : : : : : : : : : : : : :
544 LysAlaLeuPheGluAsnLysAspAsnValLysAspTrpAsnLysAlaTy 560
      : : : : : : : : : : : : : : : : : : : : : : : : :
1029 TGAGAGGAGGCTGGATCCCTCCGGGGGATATATACAGTTGG 1071
      : : : : : : : : : : : : : : : : : : : : : : : : :
560 rGluTy.....LeuPheArgAlaGluGlySerAspTrp 571
      : : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: pir2:AC1970
seq_documentation_block:
hypoetical protein alr1310 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AC1970
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC1970
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-744 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073267.1; PID:gl7130657; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
```

826 TCGGAAGTCTGCTTCCCTCAGAGCTGAAGCAGACAGTGAAGGAGGCTCTA 875
::: |||||::: |||||
438 AlaThrAlaThrIleProAlaGluGlnLeuHisSerGly..... 450

876 CITACGGACTTCGGAGTGGCCACCAAGAGCTTGAGGATATGAGAG 925
|||||::: |||||::: |||||
451SerTrpVal...AspGlySerPheThrThrTrpIleG 462

926 AGGACGAAGGAACGACAACTTAATATGCTCTCTACAAATATGAGGGC 975
|||
462 LysAspPro..... 464

976 GAACCTCGCCCTTTTACCGCAGACACGACGATGAGGGATGGAGCCCT 1025
|||||::: |||||
465AlaLysAsn.....ArgAlaTrpAspTrpLe 473

1026 CCTCAGAGAGGAGCTG..... 1041
||| |||||
473 uThrGluAlaArgIleMetLeuAlaAsnHisProGluAlaThrGluAla 490

1042CATCCCTCCGGGGGATATAT.....AAGATTGG 1071
::: |||||::: |||||
490 snAsnProGluAlaTrpGluAlaLeuTyrAlaAlaGluGlySerAspTrp 506

seq_name: pir1:B69553
seq_documentation_block:
methanol dehydrogenase regulatory protein (moxR) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: B69553
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glocke, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390: 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343
A:Accession: B69553
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-324 <LE>
A:Cross-references: GB:AE001108; GB:AE000782; NID:92589431; PIDN:AB91247.1; PID:9265067
C:Superfamily: methanol dehydrogenase regulatory protein

alignment_scores:
Quality: 120.50 Length: 327
Ratio: 0.748 Gaps: 14
Percent Similarity: 49.235 Percent Identity: 21.713

alignment_block:
US-09-886-400-3 x B69553 ..
Align seg 1/1 to: B69553 from: 1 to: 324

1 TTGAGAGCGCTGCTTTTCACGGCAACCTCCAGTATGCCGAATCCCA... 48
||| ||| ::: |||||::: |||
35 LeuAlaAlaLeuThrAsnGlyAsnIleLeuPheGluAspTyrProGlu 51

49AAGAGGGAATCCCAAGGTCATAGAGAGGACATACATCCAG 91
|||||::: |||||::: |||||
51 yLeuGlyLysThrLeuLeuAlaLysValPheAlaArgVal..... 64

92 TCATCGAGACACTGATTAAAGAAGAAATTCCTTTGGGCTCAACATAACG 141
|||
65IleGlyAlaAspTyrArgVal 72

142 GCCTATACCTTAAGTTCCTCCGAGGATATATAGACCTCGTTAAGG 191
::: |||||::: |||||
73 GlnPheThrProAspLeuLeuProSerAspIleIle..... 84

192 GGGCATC.....GGCAGTGCCTGATAGATAATC...GGAACGA 229
|||||::: |||||
85 .GlyValLysIleLeuTrpArgGlyAspArgPheGluPheValLysGlyProI 101

230 GCTACAGCAGCCCAATACCTC.....CCCTC 255
::: |||||::: |||||
101 lePheThrAsnValLeuLeuAlaAspGluIleAsnArgSerProProLys 117

256 CTCCTCGCTTACGAGAGTAGAAGCAGACAGTTCAGAGAGATAGGAAGTTAA 305
::: |||||::: |||||
118 ThrGlnAlaAlaLeuLeuGluAlaMetGluGluLysGlnIleThrValGlu 134

306 GGAAGAGCTCTTCGAGCTTTCCTCAAGGGATTCTGCTGCCAGAGCTCG 355
::: |||||::: |||||
134 uGlyGluThrPheSerLeuSerMetProPhePheValLeu.....Ala 149

356 CCTATGACCCGATATATCCCTGCCATCTGCAAGGACACAGGTTATCAGTAT 405
::: |||||
149 hrGlnAsnProIle..... 153

406 CTATTCGGCGGAGGGGAGGATGCTTTTCTCAGTCACTCACTCACTCGGC 455
::: ||||| ||| ::: |||||::: |||||
154 ...GluGlnGluGlyThrTyProLeuProGluAlaGlnMetAspArgPh 169

456 GATAAAGCCAAATAAAGCGCTCTATCCACCTTATAAGGCCCAAGAGG 505
::: |||||::: |||||
169 eMetLeuArgMetArgProGlyTyProGluSerIleGluGluGluMetG 186

506 AAAAGCGCTTTAGGTACATCAGCTAT..... 531
|| ||| |||||
186 luIleLeuArgArgIleSerTyrPheArgLysAspProThrGluAsp 202

532CTCCTGCTCTCAGGAGGCTTAGGAAGGCGAT 563
||| |||||::: |||||
203 ValGluProValValSerLeuGluThrPheArgGluGlnAspAlaVa 219

564 AAGCTCGCTTTTGAAGTGAAGTGAACCTTAAGCAGCTCAAGACATCG 613
::: |||||::: |||||
219 lGluAlaValTyrValAspLysSerIleLeuLysTyrIleSerGluLeu 236

614 AAGCGCTACCCCTGGTGGCGGTGAACAGCGCTTAATCTCGTGGCATC 663
::: |||||
236 alArgAla.....ThrArgGluHisGluLeuValGluLeuGlySer 249

664GGAAGCTCTCTTATGAATCTTAAAGAGTGGCAGCTG 704
||| ||| |||||
250 SerProArgGlyLeuAlaLeuLeuLysLeuAlaArgAlaLeuAlaVa 266

705 GATAGAGGACAAGGACAACATCTT.....CTATACGGCACCGATA 745
::: |||||::: |||||
266 lMetAspGlyArgAspPheValIleProAspValLysArgValAlaV 283

746 TAGAGTTCATTCGCTATAGGACATTCGAGGCTACAGATAGTGTGAG 795
::: |||||::: |||||
283 alGluAlaLeuAlaHisArgValIleLeuLysPheGluTyrAlaValGlu 299

796 GGATATTATTA.....GAGTWTATACGAGCTCACTCGGAAGTGTGCT 839
||||| |||||::: |||||
300 GlyLeuArgAlaGluGluValValGluGluIleLeuAsnSerValArgVa 316
::: |||||
840 TCCCTCAGAGCTGAAGCAGCAGTGGAGGGAG 870
||| |||||::: |||||
316 lPro.....LysTyrGluAlaGlnGlu 323

seq_name: pir2:C75120
seq_documentation_block:
hypothetical protein PAB1857 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: C75120
R:Anonymous, Genoscope
Submitted to the EMBL Data Library, July 1999


```

40  GAAATCCCAAGAGCGCAATCCCAAGAGTGCATATAGAGAAGCCATACATCC 89
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71  GluTyrlIleLysAlaIleupheAsnArgTyMetGluArgLysIle.... 85
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90  AGTCATCGAGACACTGATTAAAGAGAAGAAATTCCTTTGGGCTCAACATAA 139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86  .....LysLeuMetLysGluAspLeuLysLysAlaAspGlyLysL 99
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
140  CCGGGCTATACCTTAAAGTTCCTCCCGAAG..... 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99  euArgAsnAlaIleGluPheMetIleLysTyPheGluAspValTyArgL 115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169  .....GATATATATAGACCTCGCTTTAAAGGGGGCAT 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116  TyTrpSerLysIleAsnGlyAspIleIleGlyArgPheLysGlnLeuG 132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198  CCGGAGTAGCTAGATAGAGATAATCGGAACGAGCTACACCGACGCAATAC 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132  nAspGluGlyPheValGluIleIleThrSerAlaIleThrHisGlyTyrL 149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
248  TCCCTCCTCCCGGCTTAGCAGA...GTAGAAGCACAAAGTTCAGAGAGAT 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
149  euProLeuLysGlyArgAspGluAlaIleAspAlaGlnIleLeuThrGly 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295  AGGAAGTTAAGAGAGAGACTCTTCGAGCTTCTCCAAAGGATCTCGGCT 344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
166  IleArgValTyGluLysTyPheGlyLysLysProArgGlyIleIrpLe 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345  GCACAGCTCGGCTATGACCCGATA..... 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182  uproGluCysAlaTyArgProAspGlyLeuTrpLysSerProSerThrG 199
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
370  .....ATCCCTGCATACCTGAAGGCAACGGT 396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
199  lyGluIleLysTrpArgTyGlyIleGluHisPheLeuLysLysTyArgly 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
397  TATGATATCTATTCGCCGAC 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
216  LeuGluIlePhePheValGlu 222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:06:18 ; Search time 13.44 Seconds

	Search time	18.74 seconds
(without alignments)		
1048.654 Million cell updates/sec		

Title: US-09

```

perfect score: 1877
Sequence: 1 LRALVFHGNLQVAEIPKSEI.....RRLDAFRAIYNDRGNGEPP 364

```

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Minimum DB seq	length: 200000000
Maximum DB seq	length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 s

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	161.5	8.6	633	1	AMVA_PVRHO
2	157.5	8.4	685	1	AMVL_DICHT
3	145.5	7.8	655	1	AMVA_PYRAB
4	141.5	7.5	467	1	AMVA_METJA
5	141.5	7.5	648	1	AMVA_PRFU
6	121.5	6.5	653	1	MAIQ_PYRKO
7	121.5	6.5	659	1	MAIQ_THELI
8	112	6.0	311	1	MDH_BACSU
9	103.5	5.5	314	1	MDH_BACHD
10	101.5	5.4	705	1	RNR_AQUAE
11	100.5	5.4	321	1	MDH_BACTC
12	99.5	5.3	471	1	UDPG_PVRPY
13	98	5.2	474	1	NOEK_RHISN
14	97.5	5.2	813	1	LON2_ORBUB
15	96	5.1	312	1	MDH_BACIS
16	96	5.1	1067	1	LONM_SCHPO
17	93.5	5.0	349	1	ID12_SYNV3
18	93.5	5.0	1018	1	SY1_ARCFV
19	93	5.0	1570	1	P3_K1_DICDI
20	92	4.9	428	1	Y047_METJA
21	91.5	4.9	467	1	UDPG_MUSAC
22	91	4.8	852	1	CLPB_CORGL
23	91	4.8	871	1	SYA_AQUPY
24	91	4.8	871	1	SYA_AQUPY
25	90.5	4.8	476	1	BIRC_HUMAN
26	90	4.8	476	1	UDPG_SOLTU
27	90	4.8	387	1	HEMZ_SYNY3
28	90	4.8	630	1	Y235_METJA
29	89	4.7	658	1	RPE_BUCAP
30	89	4.7	344	1	CHBP_THEMEA
31	89	4.7	593	1	MTHS_YEAST
32	89	4.7	669	1	COTE_HUMAN
33	88.5	4.7	223	1	GTAL_RABIT
34	88.5	4.7	473	1	UDPG_ORVU

[illegible]

ALIGNMENTS

```

RESULT 1
AMTA_PVRHO STANDARD; PRT; 633 AA.
ID AMYA_PVRHO
AC 057932;
AC 30-MAY-2000 (Rel. 39, Created)
AC 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DT Alpha-amylase (EC 3.2.1.1).
DE AMYA OR PH0193.
GN Pyrococcus horikoshii.
OS Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OC NCBI_TaxID=53953;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=OT3;
RC MEDLINE=98344137; PubMed=9579194;
RX Kavarabayasi Y., Savada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shiizuwa H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -1- PATHWAY: POLYSACCHARIDE DEGRADATION.
CC -1- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF000001; BAA29262.1;
CC InterPro; IPR004300; Glyco_hydro_57.
CC Pfam; PF03065; Glyco_hydro_57; 1.
CC GlycoStase; Glycosidase; Carbohydrate metabolism; Complete proteome.
CC SEQUENCE 633 AA; 75017 MW; 90CBBCB2FE191501 CRC64;

```

```
Query Match      8.6%; Score 161.5; DB 1; Length 633;
Best Local Similarity 23.4%; Pred. No. 4.5e-05;
Matches 97; Conservative 60; Mismatches 130; Indels 127; Gaps 25;
```

```
y      8 GNLQVAEPKSEIPKVIEKAYIVPIETLKEIP---FGLNITGYTLKFYPK---DIIDL 61
|| :| ||| :||| ||| ||| :||| :||| :||| :|||
b     17 GNFEWI-----IKRAYEKAYRPFLET--EYPPNMKVAVHISGVLVLEWRARPEYIDL 68
||||| :||| :||| :||| :||| :||| :||| :|||
y     62 VKGGTASDLIELIGTSYTHAILPLPLSRVEAQVQRDEVKEELFELS-----PK 111
:| :| :||| :||| :||| :||| :||| :|||
b     69 LKSLLKKGGVELVWAGYEYFLVLAIP-----EDRVEQIKLSKGWAKMYEAR 117
```

```

QY      112 GFWLPDLAYDPIPAIKKNGEYELFADGEAMLFAHLNSA-IKPKPLPYLHLIKAQREK 170
        ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      118 GLWLTERVWEPELVKTLRREAGIEYVILD-----DYHMSAGLSKEELFWPYYTENGGEA 171
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY      171 RFRY-----ISYLGLURELRKAIKLVE-----EGKVTLKAVKDTAEVPVVAVNTAVML 219
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      172 IVVFPIDEKLRYILPRPNNEILEYLSLADEDESKVAV-FDDGSEKFCANFGTHELYV- 229
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY      220 GIGRLPLMNPXKKVASWI-----EDKNLLYGTDFIEFTG-----VRDIAGY-R 261
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      230 -----ERGLWKEFFDRISSDDDKINLMYLS---EYLSKFPKGLVLYPLIASYPE 274
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY      262 MSVEGL-----LEVIDELNELCLPSLKHSGRLEYLRTSSNAPDKSRLRWREDCN 313
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      275 MSEWSLPARQAALFFFEIKKL-KELNLFEKYR-----IFRGIIW---KNF-LYKYPEGN 324
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY      314 ---ARUNMLSYNRMGELLALLAENSNDARGWEPLPERLDRAFLYND--WRGENG 362
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      325 YMHRKMMLMS-----KLLRN-----PTARIFVLRAQCNDAYWHGVFG 362
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

RESULT 2
AMYL_DICTH
ID AMYL_DICTH STANDARD; PRT; 685 AA.
AC P09961;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Alpha-amylase 1 (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase).
DE AMYA.
GS Dictyoglomus thermophilum.
OS Bacteria; Dictyoglomus group; Dictyoglomus.
OX NCBI_TaxID=14;
RN [1]
RP STRAIN=H-6-12;
RX FUKUSUMI S., Kaminouchi S., Horinouchi S., Beppu T.;
FU MEDLINE=88225097; PubMed=2453362;
RT "Cloning and nucleotide sequence of a heat-stable amylase gene from
   an anaerobic thermophile, Dictyoglomus thermophilum.";
RL Eur. J. Biochem. 174;15-21(1988).
CC -!- FUNCTION: THIS AMYLASE IS A HIGHLY LIQUEFYING-TYPE: OLIGOMERS
CC APPEARED AT THE BEGINNING OF INCUBATION, FOLLOWED BY A GRADED
CC DECREASE IN THE AMOUNTS OF MALTOTRIOSE, MALTOSE AND GLUCOSE IN
CC PROLONGED INCUBATION. IT IS HIGHLY HEAT-STABLE; THE OPTIMUM
CC TEMPERATURE FOR ITS ACTIVITY WAS FOUND TO BE AROUND 90 DEGREES
CC CELSIUS, BUT DECREASED TO ONLY 70% ACTIVITY AFTER 1 HOUR. NO
CC DECREASE OF ACTIVITY WAS OBSERVED WITHIN THE SAME TIME AT 80
CC DEGREES CELSIUS.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X07896; CAA30735.1; -.
DR PIR; S00628; ALDAYT.
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03063; Glyco_hydro_57; 1.
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Multigene family.
FT INIT_MET 0
SQ SEQUENCE 685 AA; 81060 MW; C10941C8A508C404 CRC64;

```

DB 28 EKAYRPFLEIL--EYPPNKMVAIHISGILVLEENRDPYIDILKSLVRKGQVEIYVAGF 85
QY 79 THAILPLPLSRVQAQVQDRREVKEELFELSPKGFWLPAYDPIIPAILKONGEYELFA 138
DB 86 YEPVLAAPKEDRLEQIYLLKWKAKKI-GYDAKGLWTERWQPELVKTLREAGIEVVVV 144
QY 139 DGEAMFSAHLNSA-IKPIKPLYPHLIKAQRE-----KRFY-----ISYLL 179
DB 145 D-----DYHMSAGLSKDLQFWPYTDEGSEVITVPIDKRLYLIPFRPDKVISYLH 198
QY 180 GLRELKRAIKLVFEKGVTLKAVKDIEAVPVAVNTAVMLGIGRLPLMNPCKVASWTECK 239
DB 199 SLASEDESKVAFH-----DDGEKFGIW-----PM-----IYEWYVEK 231
QY 240 D-----NILLVGTDIE-----FIGYRIAGY-----RMSVEG 266
DB 232 GWLREFFRVSSDEAINIMLYSEYLOKFKGLVLYPIASYFEMSENSLPAQAKLFVE- 290
QY 267 LLEVIDELNSELCLPSLKHSGRELYLTSWAPDKSLRWREDEGN---ARLNMLSYNM 323
DB 291 FVEKLEKLN-----MFERYRVFVGGIW---KNP-FKYYPEANYMHKMLMLS--- 334
QY 324 RGEALLAENSARDGWEPLPERRLDAPRAIYND---WRGENG 362
DB 335 ----RLLRDNPSARF-----VLRQCNDAWYHGVFG 362
RESULT 4
AMTA_PYRFU
ID AMYA_METJA STANDARD; PRT; 467 AA.
AC Q59006;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative alpha-amylase (EC 3.2.1.1).
GN M0611.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: U67601; AAB99631.1; -
CC TIGR: M16111; -
CC InterPro: IPR004300; Glyco_hydro.57.
CC Pfam: PF03065; Glyco_hydro.57; 1.
KW Hypothetical protein; Hydrolase; Glycosidase; Carbohydrate metabolism;

KW Complete proteome.
SQ SEQUENCE 467 AA; 55558 MW; 40A6B1CDD4D967E CRC64;
Query Match 7.5%; Score 141.5; DB 1; Length 467;
Best Local Similarity 22.8%; Pred. No. 0.0011;
Matches 86; Conservative 66; Mismatches 130; Indels 95; Gaps 24;
QY 14 EIPKSEIPKVIKAYIPVETLIK--EELIFGLN--ITG-----YTLKFLPKD-IIDLKVG 64
DB 34 KLNKEVFNKANKYIPTNELILEIDYDFKVNYSYTVFVQALEF--NDYVLDLFDK 91
QY 65 GIASDLIIITSTHAILPLPL--SRVEAQVQDRREVKEELFELSPKGFWLPAYDPI 123
DB 92 LVKTVNGVELIAETVYHSLTSLFETEDFEIDEMHRKMYKEITCFKAKVPRNTELIYNNR 151
QY 124 IPAILKONGEYELFADG-EAMLFSAHLNSAIKPIKPLYPHLIKAQREKRYYS---YLL 179
DB 152 TAKIAKDLGKRAITEGIEKIL-----GWRSPNYLYOSPODKMI 190
QY 180 GLRELKRAIKLVFEKGVTLKAVKDIEAVP-----VWAVNTAVMLGIGRLPLMNPCKV 232
DB 191 LLRNYSDDIGR-----FSARDWDQVPLTADKYAIWLASTPGEVINI---YMDYET 241
QY 233 AS--WIEDKDNILLYGTDI-EFIGYR--DIAGY-RMSVEGLLEVIDELNSELCLPSLKH 286
DB 242 GEHWKE-----TGIFELRYLPETIAKHEHLEVNVYVSDLEPR----- 283
QY 287 SGRELYLR---TSSWA-PDKSLRWREDEGNARLNMLSYNMRGELA--LLAENSARDGWE 341
DB 284 --GEIYVHEFATISWADTERDYSAWL---GN-KMQRISFEKLKDIGKFIKENS 333
QY 342 LPERRLDAPRAIYNDWR 358
DB 334 ---KKLNKFDIEIKMYK 347
RESULT 5
AMTA_PYRFU
ID AMYA_PYRFU STANDARD; PRT; 648 AA.
AC P49067;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-amylase (EC 3.2.1.1).
GN AMYA OR PF0272.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94043280; PubMed=8226590;
RA Laderman K.A., Asada K., Uemori T., Mukai H., Taguchi Y., Kato I.,
RA Anfinsen C.B.;
RT "Alpha-amylase from the hyperthermophilic archaeobacterium Pyrococcus
RT furiosus. Cloning and sequencing of the gene and expression in
RT Escherichia coli.";
RL J. Biol. Chem. 268:24402-24407(1993).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=94043279; PubMed=8226989;
RA Laderman K.A., Davis B.R., Krutzsch H.C., Lewis M.S., Griko Y.V.,
RA Privalov P.L., Anfinsen C.B.;
RT "The purification and characterization of an extremely thermostable
RT alpha-amylase from the hyperthermophilic archaeobacterium Pyrococcus
RT furiosus.";
RL J. Biol. Chem. 268:24394-24401(1993).
CC -!- FUNCTION: DISPLAYS A BROAD RANGE OF SUBSTRATE SPECIFICITY, WITH
CC THE CAPACITY TO HYDROLYZE CARBOHYDRATES AS SIMPLE AS MALTOSE.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.
CC -!- SUBUNIT: HOMODIMER.

Db 235 AASLAEMVEAILKDK-----KRVLTIAYL-----EGEYGED 267
QY 175 -----ISYLGELRELKRAIKLVPEGVKVLKAVKDIKAV 207
Db 268 IYGVPTILGGDIEKIVELDLTDEKATFAKIESV 304

RESULT 10
RNR_AQUAE STANDARD; PRT; 705 AA.
ID RNR_AQUAE
AC 067834;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonuclease R (EC 3.1.-.-) (RNase R) (VacB protein homolog).
GN RNR OR VACB OR AQ_2046.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V., Swanson R.V.,
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -1- FUNCTION: 3'-5'EXORIBONUCLEASE THAT PARTICIPATES IN AN ESSENTIAL
CELL FUNCTION. ACTS NONSPECIFICALLY ON POLY(A), POLY(U) AND
CC RIBOSOMAL RNAs (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SI MOTIF DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE007769; AAC07792.1;
DR InterPro: IPR002059; Cold_shock.
DR InterPro: IPR001900; Ribonuclease_II.
DR InterPro: IPR003029; SI.
DR Pfam: PF00773; RNB; 1.
DR Pfam: PF00575; SI; 1.
DR SMART: SM00357; CSP; 1.
DR SMART: SM00316; SI; 2.
DR PROSITE: PS01175; RIBONUCLEASE_II; 1.
DR PROSITE: PS01128; SI; 1.
KW Hydrolase; Exonuclease; Nuclease; RNA-binding; Complete proteome.
FT DOMAIN 515 696 SI MOTIF.
SQ SEQUENCE 705 AA; 81537 MW; 2C08587229D1B05E CRC64;

Query Match 5.4%; Score 101.5; DB 1; Length 705;
Best Local Similarity 16.6%; Pred. No. 2.6;
Matches 77; Conservative 79; Mismatches 114; Indels 195; Gaps 22;
QY 5 VFHGNLOYAEIP---KSE--IPKVIKAYIPVETLIKSE-----IPFGLNITGTLKF 53
Db 100 VFHGVVRAKVTETFGKKEVRIKVLKRAKDIKAVKVFEDQCYVVPDEN-AHRILL 158
QY 54 LPKDIIDLKVGKIA-----SDLIBIG-----TSVT 79
Db 159 SKKDCQCKLKEGEVVLKVTQPTTKSPARGKVIENLGNPKKFAIDVIIRKYNLTSP 218
QY 80 HAIL---PLLPLSRVEAQVQRDRVKEEL-----FELSPKGFV----- 114
Db 219 EKVIREVAIPETIEEIKRRRLRQLCTIDPEKAGDFDDAVALETPEGVYKLVH 278

Db 235 ASLTEMVEAILKQRRVLTIALEYGEYGIYGVPTIVGGNGLEQIELELTDERA 294
QY 148 HLNSAKPK 157
Db 295 QLNKSVSVK 304

RESULT 9
MDH_BACHD STANDARD; PRT; 314 AA.
ID MDH_BACHD
AC Q9K849;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
GN MDH OR CITH OR BH3158.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AP001517; BAB06877.1;
DR InterPro: IPR001557; L_LDH.
DR InterPro: IPR001252; MDH_active.
DR InterPro: IPR002005; NAD_binding.
DR InterPro: IPR001236; LDH.
DR Pfam: PF00056; LDH; 1.
DR Pfam: PF02865; LDH_C; 1.
DR PRINTS: PR00086; LLDHRCGNASE.
DR PROSITE: PS00068; MDH; FALSE_NEG.
KW Oxidoreductase; Tricarboxylic acid cycle; NAD; Complete proteome.
FT ACT_SITE 153 153 PROTON-RELAY (BY SIMILARITY).
FT BINDING 156 156 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).
FT ACT_SITE 180 180 PROTON-RELAY (BY SIMILARITY).
SQ SEQUENCE 314 AA; 33680 MW; 02D132F11E3B8E34 CRC64;

Query Match 5.5%; Score 103.5; DB 1; Length 314;
Best Local Similarity 23.0%; Pred. No. 0.61;
Matches 50; Conservative 32; Mismatches 74; Indels 61; Gaps 9;
QY 10 LOYAEIPKSEIPK-----VIEKAYIPVETLIKSEIPGL-NINGYTLKPLPKDID 60
Db 130 MITYVYKSGFPKPNVIGSGVLDTPAR---FRTEVAGELNLSVEDITGFVL----- 177
QY 61 LVKGGIASDLIEITGTSYTHAT--LPLPLSRVEAQVQRDRVKEELFELSPKG--FWLP 116
Db 178 ---GGHDDMVPLIRYSAGGIPLEKLLPQERDAIVERTKGGIEVGLIGNSSAYAP 234
QY 117 ELAYDPIIPAILKNGEYVLFADGEMFLFSAHNSAIKPIKPLPLPHLKAGREKRFY-- 174

QY 109 SPKG-----FWPELAYDPIIPAILKONGYEVYLFADGEAMLFSAHLNS--AIKPIKPLYP 161
 DB 179 PSKGQTKGDKWYPP--GHGDVFFSLKNSGKLDLLLSQGREYVFIA NSDNLGAVVDKILH- 236
 QY 162 HLKAKQREKRFVIVSLL-----GLRELKAKILVFEKGVILKAV-----KDIE 205
 DB 237 HLLQKNE-----YCEVTPKTLADVKGGTLISYERGVQLLETAQVPODHOVNEFKSIE 289
 QY 206 AVFV-----WVAVNT-----AVMLGIGRLPLMNPVKV 232
 DB 290 KFKIFNTNLLWNLNAIKRLVADALKMEIIP--NPKEV 326

RESULT 13
 NOEK_RHISN STANDARD; PRT: 474 AA.
 AC P55356;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Phosphomannomutase (EC 5.4.2.8) (PMM).
 GN NOEK OR Y4A1.
 OS Rhizobium sp. (strain NGR234).
 OG Plasmid sym pNGR234a.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97305956; PubMed=9163424;
 RA Freiberg C.A., Feilay R., Bairoch A., Broughton W.J., Rosenthal A.,
 RT Perret X.;
 RT "Molecular basis of symbiosis between Rhizobium and legumes.";
 RL Nature 387:394-401(1997).
 CC -1- CATALYTIC ACTIVITY: D-mannose 1-phosphate -> D-mannose 6-phosphate.
 CC -1- PATHWAY: BIOSYNTHESIS OF THE FUCOSE MOIETY OF THE NOD FACTOR.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOHEXOSE MUTASES FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE000064; AAB91606.1;
 CC InterPro: IPR001485; PGM_PMM.
 CC Pfam: PF00408; PGM_PMM.1
 CC Pfam: PF02878; PGM_PMM.1; 1
 CC Pfam: PF02879; PGM_PMM.1; 1
 CC Pfam: PF02880; PGM_PMM.1; 1
 CC PROSITE: PS00710; PGM_PMM; 1.
 CC Isomerase; Phosphorylation; Modulation; Plasmid.
 FT ACT_SITE 102 102
 FT (BY SIMILARITY).
 SQ SEQUENCE 474 AA; 49670 MW; 6965310CEA96A22B CRC64;

Query Match 5.2%; Score 98; DB 1; Length 474;
 Best Local Similarity 20.1%; Pred. No. 2.9;
 Matches 79; Conservative 57; Mismatches 133; Indels 124; Gaps 15;

QY 15 IPKSEIPKVIKA-----YIPVETILKEIIPF 42
 DB 114 LPDGEINKADSOITALEAQLSADADATRVCECGADHSEATDFIQRYELLPSGLK 173
 QY 43 GLNITGYTLKPLKPDHDLVKGGIASDLIEIGTSYTHAILPLPLSRVDAQVQRDEVK 102
 DB 174 GLKIGLYCHSSVARDILTTILEGCAVW-----PVGRSEVFIPVDTEAI 218
 QY 103 BELFELSPKGFWLVDYFIIPA-----ILKDNQYEVLPAD-----GEAMLF 146

DB 219 SAA-TCKMLAAWAKFAFDAIVSSDADADRELLTDTCTPLRGDLGLGLICARLEAKLIA 277
 QY 147 AHL--NSAIKPIKPL-----YPLIKAQREKRFRYISYLLGLRELKRAIKLVFE--- 193
 DB 278 TPITSSGIEAASGVEVVRTRVGSVPYIAAMTEAVARGKQRMGF-EANGGVMLGSNFSF 336
 QY 194 GKVTLKAVKDIEAV-PVWVAVNTAV-----MLGIGRLPLMNPVKVASWIEDKDNL- 243
 DB 337 GGASLPALPTRDCVLPITIAALHMAVEAKTPLSGIAMHRLPVALSGRIENPPFRSALY 396
 QY 244 -----LYGTDIEFIGYRDIAGYRMSVEGLLEVIDELANSELCLPSELKHSREL 291
 DB 397 AFLKASKANVSHLFSRIGRVAGTDDVQGLRLTFEG-----GRIL 435
 QY 292 YLRTSSWAPDKSLRIHRE-DEGNARLNMLSYNM 323
 DB 436 HIRPSGNAPE--LRCYVEADDPDAEHLAQL 466

RESULT 14
 LON2_BORBU STANDARD; PRT: 813 AA.
 AC Q51556;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ATP-dependent protease La homolog (EC 3.4.21.-).
 GN BR0613
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Hickey E.K., Gwinn M.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Richardson D.,
 RA Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kleravage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Uitterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 RT burgdorferi.";
 RL Nature 390:580-586(1997).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16; ALSO KNOWN AS THE LON
 CC FAMILY OF ATP-DEPENDENT PROTEASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE001162; AAC66962.1;
 CC TIGR: BB0613;
 CC InterPro: IPR003593; AAA.
 CC InterPro: IPR003599; AAA_subfam.
 CC InterPro: IPR001687; ATP_GTP_A.
 CC InterPro: IPR003111; LON.
 CC InterPro: IPR001984; Lon_endopep.
 CC Pfam: PF00004; AAA; 1.
 CC Pfam: PF02190; LON; 1.
 CC PRINTS: PR00830; ENDOLAPTASE.
 CC SMART: SMC0382; AAA; 1.
 CC SMART: SMC0464; LON; 1.
 CC PROSITE: PS01046; LON_SER; 1.
 KW Hydrolase; Serine protease; ATP-binding; Complete proteome.
 FT NP_BIND 369 376
 FT BY SIMILARITY.
 FT ACT_SITE 719 719

```
SQ SEQUENCE 813 AA; 92312 MW; 5129AA1498C5D0F6 CRC64;

Query Match
Best Local Similarity 5.2%; Score 97.5; DB 1; Length 813;
Matches 72; Conservative 52; Mismatches 113; Indels 95; Gaps 18;

QY 12 YAEIPKSE-----IPKVEIKAYIPVIETLKEEIPFGINTGYT----- 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 507 YSIEKLEIAKIFLPSIISKIFDKVIKRIEDDVIFNL-IRNTMESGVGLKRVITNL 565

QY 51 LKPLPKDII-----DLVKGGI--ASDLTEIIGTSYTH-----AILPLPLSRVQAQVR 97
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 566 IRLVRELLVEYSKDQIIKNGEYSPSLIHGNSLFTHPDIPGIIYKIINNNYNYV-- 623

QY 98 DREVKEELFELSPKGF-----W-----LP-----ELAYDPIIPAILKDNQYEV 135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 624 DTEEDNLDLIKSSSGFVYGLAWNTYGGIVLPVEATKFKKGGDIITGSLGAIMKESAQ-- 681

QY 136 LFADGEAMLFSAHLSAIPKIPLYPHLIKAQREKFRFYISYLLG-----LRELKAIK 189
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 682 -LAYSIVKYSSKLNFDVKESPEIHLHPPEGATPKDGSAGITATASILSKKVPFLD 740

QY 190 LVFEGKVTILKAYKDIKAVPW---VAVTAVMLGIGR--LPLMNEKKYASVIED-KONIL 243
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 741 LAMTGEVTLKGF-----VLPVGGIKKVKLAAYRNGISKVILPKDNKKDYKLPVEVKDNI- 795

QY 244 LYGTDIEFIQYRDIAGYRMSVSGLELVEIDELN 275
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 796 ----DVKE-----VSSLEEVFDYLN 811

RESULT 15
MDH_BACIS
ID MDH_BACIS STANDARD; PRT; 312 AA.
AC Q59202;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
GN MDH.
OS Bacillus israeli.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=42189;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96276423; PubMed=8694770;
RA Wynne S.A., Nicholls D.J., Scawen M.D., Sundaram T.K.;
RT "Tetrameric malate dehydrogenase from a thermophilic Bacillus:
RT cloning, sequence and overexpression of the gene encoding the enzyme
RT and isolation and characterization of the recombinant enzyme.";
RL Biochem J. 317:235-245(1996)
CC -|- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC -|- SUBUNIT: HOMOTETRAMER.
CC -|- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X90527; CAA62129.1; -
DR HSSP; Q27743; ICET.
DR InterPro; IPR001557; L_LDH.
DR InterPro; IPR001252; MDH_actsite.
DR InterPro; IPR000205; NAD_binding.
DR InterPro; IPR001236; ldh.
DR Pfam; PF00056; ldh; 1.
DR Pfam; PF02866; ldh_C; 1.
```

```
DR PRINTS: PR00086; LLDH5RGNAE.
DR PROSITE; PS00068; MDH; FALSE_NEG.
KW Oxidoreductase; Tricarboxylic acid cycle; NAD.
FT ACT_SITE 153 153 PROTON-RELAY (BY SIMILARITY).
FT BINDING 156 156 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).
FT ACT_SITE 180 180 PROTON-RELAY (BY SIMILARITY).
SQ SEQUENCE 312 AA; 33648 MW; 64CE3BFF1B3C02D2 CRC64;
```

```
Query Match
Best Local Similarity 5.1%; Score 96; DB 1; Length 312;
Matches 36; Conservative 23; Mismatches 48; Indels 26; Gaps 6;

QY 10 LOVAEIPKSEIPK-----VIEKAYIPVIETLKEEIPFGINTGYTLKFLPKDIIDL 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 MIYVYKESGFFKHRVIGSGVLDIAR---FRTFVAQE---LNL-----VKDITGF 175

QY 62 VKGGIASDLIEIIGTSYTHAI--LPLPLSRVQAQVRDREVKEELFELSPKG--FWLPE 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 176 VLGGHGDMMVPLRVYSYAGGIFLEKLIPIKERLEAIVERTKGGGEIVNLLNGSAYAPA 235

QY 118 LAYDPIIPAILKD 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 236 ASLVENVEAIVKD 248
```

Search completed: June 2, 2002, 18:09:55
Job time: 217 sec

